(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 August 2004 (12.08.2004)

PCT

(10) International Publication Number
WO 2004/067743 A1

(51) International Patent Classification[7]: C12N 15/11, 15/63, C07H 21/04, A01H 5/00

(21) International Application Number:
PCT/US2003/041269

(22) International Filing Date:
24 December 2003 (24.12.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/441,377	21 January 2003 (21.01.2003)	US
60/441,381	21 January 2003 (21.01.2003)	US
60/441,392	21 January 2003 (21.01.2003)	US
607441,405	21 January 2003 (21.01.2003)	US
60/441,447	21 January 2003 (21.01.2003)	US
60/441,502	21 January 2003 (21.01.2003)	US
10/609,019	26 June 2003 (26.06.2003)	US

(71) Applicants *(for all designated States except US)*: **TRANSGENRX, INC.** [US/US]; 1755 Wittington Place Drive, Suite 300, Dallas, TX 75234 (US). **THE BOARD OF SUPERVISORS OF LOUISIANA STATE UNIVERSITY AND AGRICULTURAL AND MECHANICAL COLLEGE** [US/US]; P.O. Box 25055, Baton Rouge, LA 70894 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **COOPER, Richard, K.** [US/US]; 111 Pecan Meadow Drive, Baton Rouge, LA 70810 (US). **FIORETTI, William, C.** [US/US]; 2225 Lakeridge Road Drive, Grapevine, TX 76051 (US). **CADD, Gary, G.** [US/US]; 501 Turner Road, Apartment 1111, Grapevine, TX 76051 (US).

(74) Agent: **PRATT, John, S.**; Kilpatrick Stockton LLP, Suite 2800, 1100 Peachtree Street, Atlanta, GA 30309 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: VACCINE PRODUCTION USING TRANSPOSON BASED VECTORS

(57) Abstract: The present invention provides novel methods of making proteins, multimeric proteins and antibodies using transposon-based vectors in transgenic individuals that are desirable in producing vaccines. The present invention further provides a method for providing protection against disease by vaccinating an individual with the proteins, multimeric proteins and/or antibodies made in the transgenic individuals of the present invention.

WO 2004/067743 A1

VACCINE PRODUCTION USING TRANSPOSON BASED VECTORS

The U.S. Government has certain rights in this invention. The development of this invention was partially funded by the United States Government under a HATCH grant from the United States Department of Agriculture, partially funded by the United States Government with Formula 1433 funds from the United States Department of Agriculture and partially funded by the United States Government under contract DAAD 19-02016 awarded by the Army.

FIELD OF THE INVENTION

The present invention relates generally to production of proteins, polypeptides and peptides in a transgenic individual, wherein genes encoding the proteins, polypeptides and peptides are operably-linked to signal sequences, or portions of signal sequences. The proteins, polypeptides and peptides are useful as vaccines.

BACKGROUND OF THE INVENTION

Diseases caused by numerous biological organisms result in widespread morbidity and mortality throughout the world. Infectious disease is rampant in many locations, and unfortunately, numerous individuals suffer from more than one disease. Immunization of animals and humans with compositions comprising vaccines capable of generating an immune response is one of the only means of preventing or ameliorating the severity of disease.

Numerous diseases threaten not only humans but also animals. Influenza epidemics kill millions each year. Suppy of influenza vaccine is often inadequate for

the needs of numerous individuals. Further, the influenza vaccine is often designed on the best guesses of infectious disease experts concerning the expected antigens and strains for any given year. Different governments and the Centers for Disease Control face a major challenge in redesigning a vaccine and rapidly producing it to supply those in need.

Recent bioterrorism, for example the anthrax attacks, demonstrates the need for a stockpile of antidote. Bioterrorism against animals may also threaten the food supplies of nations. Engineered organisms provide a particular challenge for nations to rapidly adjust their vaccine production to attack newly discovered epitopes in these organisms.

Recent events have demonstrated the vulnerability of the U.S. to attack with biowarfare agents. For many of the agents listed as select agents by the Centers for Disease Control, vaccines are not readily available for humans or domestic livestock. Despite intense research with many of these organisms, the potential for a vaccine in the immediate future is not likely. In addition, treatment with antibiotics may or may not be effective, especially if a bacterium has been engineered to be multi-drug resistant.

Accordingly what is needed in a method for rapid and economical production of proteins, polypeptides, peptides and mixtures thereof that may be used as vaccines to prevent or ameliorate disease.

SUMMARY OF THE INVENTION

The present invention provides a new, effective and efficient method of producing proteins, polypeptides and peptides (hereinafter generally "proteins") in a transgenic individual that may be used as vaccines. The present invention provides the capability to transform the genetic machinery of individuals in order to produce vaccines. Such transformation may be germ-line transformation and/or non germ-line transformation. The present invention facilitates production of one or more protein, polypeptide and peptide for use in vaccines.

This invention provides polynucleotide cassettes containing at least one gene of interest and one or more pro polynucleotide sequences, wherein each gene of interest is operably-linked to a pro nucleotide sequence. Each of the genes of interest encodes a protein that may be used in the vaccine. One discovery of the present

invention is the use of pro portions of prepro signal sequences to facilitate appropriate processing, expression, and/or formation of proteins in an individual. Preferred pro polynucleotides are provided in SEQ ID NO:1 and SEQ ID NO:2.

Another discovery of the present invention is that cecropin prepro sequences facilitate appropriate processing, expression, and/or formation of desired proteins in an individual. Accordingly, the present invention includes polynucleotide cassettes containing at least one gene of interest operably-linked to a cecropin prepro sequence. Preferred cecropin prepro polynucleotides are provided in SEQ ID NO:3 and SEQ ID NO:4. The present invention also includes polynucleotide cassettes containing at least one gene of interest operably linked to a cecropin prepro polynucleotide, wherein pro sequences are located between the at least one gene of interest when more than one gene of interesst is present.

A non-limiting list of the diseases and organisms that may be targets of the vaccines of the present invention follows: Anthrax (*Bacillus anthracis*), *Bacillus anthracis* (anthrax), Botulism (*Clostridium botulinum* toxin), *Brucella* species (brucellosis), Brucellosis (*Brucella* species), *Burkholderia mallei* (glanders), *Burkholderia pseudomallei* (melioidosis), *Chlamydia psittaci* (psittacosis), Cholera (*Vibrio cholerae*), *Clostridium botulinum* toxin (botulism), *Clostridium perfringens* (Epsilon toxin), *Coxiella burnetii* (Q fever), *E. coli* O157:H7 (*Escherichia coli*), Emerging infectious diseases such as Nipah virus and hantavirus, Epsilon toxin of *Clostridium perfringens*, *Escherichia coli* O157:H7 (*E. coli*), Food safety threats (e.g., *Salmonella* species, *Escherichia coli* O157:H7, *Shigella*), *Francisella tularensis* (tularemia), Glanders (*Burkholderia mallei*), Melioidosis (*Burkholderia pseudomallei*), Plague (*Yersinia pestis*), Psittacosis (*Chlamydia psittaci*), Q fever (*Coxiella burnetii*), Ricin toxin from *Ricinus communis* (castor beans), *Rickettsia prowazekii* (typhus fever), *Salmonella* species (salmonellosis), *Salmonella* Typhi (typhoid fever), Salmonellosis (*Salmonella* species), *Shigella* (shigellosis), Shigellosis (*Shigella*), Smallpox (variola major), Staphylococcal enterotoxin B, Tularemia (*Francisella tularensis*), Typhoid fever (*Salmonella* Typhi), Typhus fever (*Rickettsia prowazekii*), Variola major (smallpox), *Vibrio cholerae* (cholera), Viral encephalitis (alphaviruses [e.g., Venezuelan equine encephalitis, eastern equine encephalitis, western equine encephalitis]), Viral hemorrhagic fevers (filoviruses [e.g., Ebola, Marburg] and arenaviruses [e.g., Lassa, Machupo]), Water safety threats (e.g., *Vibrio*

cholerae, Cryptosporidium parvum), and *Yersinia pestis* (plague). Other diseases and organisms that may be targets of the vaccines of the present invention are described in Harrison's Principles of Internal Medicine, (Fauci et al. eds, 14th edition, McGraw Hill, New York, USA (1998)).

These polynucleotide cassettes are administered to an individual for expression of polypeptide sequences and the formation of a desired protein or multimeric protein for use as a vaccine. Preferably, the individual is an animal from which the multimeric protein can be harvested. In one embodiment, preferred animals are egg-laying or milk-producing animals.

In one embodiment, the egg-laying transgenic animal is an avian. The method of the present invention may be used in avians including Ratites, Psittaciformes, Falconiformes, Piciformes, Strigiformes, Passeriformes, Coraciformes, Ralliformes, Cuculiformes, Columbiformes, Galliformes, Anseriformes, and Herodiones. Preferably, the egg-laying transgenic animal is a poultry bird. More preferably, the bird is a chicken, turkey, duck, goose or quail. Another preferred bird is a ratite, such as, an emu, an ostrich, a rhea, or a cassowary. Other preferred birds are partridge, pheasant, kiwi, parrot, parakeet, macaw, falcon, eagle, hawk, pigeon, cockatoo, song birds, jay bird, blackbird, finch, warbler, canary, toucan, mynah, or sparrow.

In some embodiments, the polynucleotide cassettes are located within transposon-based vectors that allow for incorporation of the cassettes into the DNA of the individual. The transposon-based vectors of the present invention include a transposase, operably-linked to a first promoter, and a coding sequence for a protein or peptide of interest operably-linked to a second promoter, wherein the coding sequence for the protein or peptide of interest and its operably-linked promoter are flanked by transposase insertion sequences recognized by the transposase. The transposon-based vector also includes the following characteristics: a) one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:5) at the 3' end of the first promoter to enhance expression of the transposase; b) modifications of the codons for the first several N-terminal amino acids of the transposase, wherein the nucleotide at the third base position of each codon is changed to an A or a T without changing the corresponding amino acid; c) addition of one or more stop codons to enhance the termination of transposase synthesis; and/or, d) addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of

the transposase gene. In some embodiments, the effective polyA sequence is an avian optimized polyA sequence.

In one embodiment, the transposon-based vector comprises an avian optimized polyA sequence and does not comprise a modified Kozak sequence comprising ACCATG (SEQ ID NO:5). One example of such a transposon-based vector is the pTnMCS vector (SEQ ID NO:56). In another embodiment the transposon-based vector comprises a) one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:5) at the 3' end of the first promoter to enhance expression of the transposase; b) modifications of the codons for the first several N-terminal amino acids of the transposase, wherein the third base of each codon was changed to an A or a T without changing the corresponding amino acid; c) addition of one or more stop codons to enhance the termination of transposase synthesis; and, d) addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of the transposase gene. One example of such a transposon-based vector is the pTnMod vector (SEQ ID NO:57).

Accordingly, it is an object of the present invention to provide a new transposon-based method of producing proteins useful as vaccines.

It is another object of the present invention to provide a new transposon-based method of producing proteins useful as vaccines, wherein the proteins are antibodies.

It is another object of the present invention to provide a new transposon-based method of producing proteins useful as vaccines, wherein the proteins are multimeric proteins.

It is another object of the present invention to provide a new transposon-based method of producing proteins useful as vaccines, wherein the proteins are monoclonal antibodies.

Yet another object of the present invention is to vaccinate individuals, particularly animals and humans, with the vaccines of the present invention in order to prevent or ameliorate disease.

Yet another object of the present invention is to vaccinate individuals, particularly animals and humans, with a vaccine comprising a combination of the proteins or multimeric proteins of the present invention, together with antibodies of the present invention, in order to prevent or ameliorate disease.

Another object of the present invention is to provide proteins, multimeric proteins and antibodies that are useful in the preparation of a medicament that may be used as a vaccine.

An advantage of the present invention is that the proteins are produced more efficiently, in greater quantities and less expensively than taught by prior art methods.

These and other objects, features and advantages of the present invention will become apparent after a review of the following detailed description of the disclosed embodiments and claims.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 depicts schematically a polynucleotide cassette containing two genes of interest operably-linked to two pro polynucleotides, wherein the first pro polynucleotide is a part of a prepro polynucleotide. "Prom" indicates promoter.

Figure 2 depicts schematically a polynucleotide cassette containing polynucleotides encoding for a heavy chain and a light chain of an antibody. "Oval prom" indicates an ovalbumin promoter. The polynucleotide cassette contains pro and prepro sequences and is flanked by insertion sequences (IS) recognized by a transposase.

Figure 3 depicts schematically a polynucleotide cassette containing a cecropin prepro sequence operably-linked to two genes of interest. Between the genes of interest resides a cleavage site indicates by "CS."

Figure 4 depicts schematically a polynucleotide cassette containing two genes of interest, a promoter (prom), a signal sequence (SS) and a cleavage site (CS). The polynucleotide cassette is flanked by insertion sequences (IS) recognized by a transposase.

Figure 5 is a picture of a gel showing partially purified egg white derived from a transgenic avian run under reducing and non-reducing conditions.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides a new, effective and efficient method of producing proteins in an individual that may be used as vaccines. Such proteins include polypeptides, peptides and mixtures thereof and are generally called "proteins" throughout the present application. One or more proteins may be made

with the present invention. Antibodies may be made with the present invention and used as vaccines, alone or in conjunction with other proteins and multimeric proteins. Multimeric proteins may be produced with the persent invention and include at least two proteins, wherein the at least two proteins may be the same of different. A multimeric protein includes associated multimeric proteins (two or more associated polypeptides) and multivalent multimeric proteins (a single polypeptide encoded by more than one gene of interest). Expression and/or formation of the multimeric protein in the individual is achieved by administering a polynucleotide cassette containing the gene or genes of interest to the individual. The polynucleotide cassette may additionally contain one or more pro sequences, prepro sequences, cecropin prepro sequences, and/or cleavage site sequences.

One or more genes of interest for incorporation into the polynucleotide cassette are chosen based on the genes that generally code for epitopes of organisms that are believed to be antigenic. Such coding sequences are generally known to one of ordinary skill in the art and can be obtained through GenBank and the published scientific and patent literature. It is within the scope of the present invention to search the genome of organisms to identify coding regios for particular proteins, polypeptides, peptides or epitopes that could serve as epitopes for vaccines. A protein may be made for use in a vaccine against any desired disease or disease producing organism using the method of the present invention. A non-limiting list of the diseases and organisms that may be targets of the vaccines of the present invention follows: Anthrax (*Bacillus anthracis*), *Bacillus anthracis* (anthrax), Botulism (*Clostridium botulinum* toxin), *Brucella* species (brucellosis), Brucellosis (*Brucella* species), *Burkholderia mallei* (glanders), *Burkholderia pseudomallei* (melioidosis), *Chlamydia psittaci* (psittacosis), Cholera (*Vibrio cholerae*), *Clostridium botulinum* toxin (botulism), *Clostridium perfringens* (Epsilon toxin), *Coxiella burnetii* (Q fever), *E. coli* O157:H7 (*Escherichia coli*), Emerging infectious diseases such as Nipah virus and hantavirus, Epsilon toxin of *Clostridium perfringens*, *Escherichia coli* O157:H7 (*E. coli*), Food safety threats (e.g., *Salmonella* species, *Escherichia coli* O157:H7, *Shigella*), *Francisella tularensis* (tularemia), Glanders (*Burkholderia mallei*), Melioidosis (*Burkholderia pseudomallei*), Plague (*Yersinia pestis*), Psittacosis (*Chlamydia psittaci*), Q fever (*Coxiella burnetii*), Ricin toxin from *Ricinus communis* (castor beans), *Rickettsia prowazekii* (typhus fever), *Salmonella* species

(salmonellosis), *Salmonella* Typhi (typhoid fever), Salmonellosis (*Salmonella* species), *Shigella* (shigellosis), Shigellosis (*Shigella*), Smallpox (variola major), Staphylococcal enterotoxin B, Tularemia (*Francisella tularensis*), Typhoid fever (*Salmonella* Typhi), Typhus fever (*Rickettsia prowazekii*), Variola major (smallpox), *Vibrio cholerae* (cholera), Viral encephalitis (alphaviruses [e.g., Venezuelan equine encephalitis, eastern equine encephalitis, western equine encephalitis]), Viral hemorrhagic fevers (filoviruses [e.g., Ebola, Marburg] and arenaviruses [e.g., Lassa, Machupo]), Water safety threats (e.g., *Vibrio cholerae, Cryptosporidium parvum*), and *Yersinia pestis* (plague). Other diseases and organisms that may be targets of the vaccines of the present invention are described in Harrison's Principles of Internal Medicine, (Fauci et al. eds, 14th edition, McGraw Hill, New York, USA (1998)).

This invention provides polynucleotide cassettes containing at least one gene of interest and one or more pro polynucleotide sequences, wherein each gene of interest is operably-linked to a pro nucleotide sequence. When one gene of interest is present in the polynucleotide cassette, a single protein is made, as opposed to a multimeric protein as described below. The present invention provides polynucleotide cassettes containing two or more genes of interest and two or more pro polynucleotide sequences, wherein each gene of interest is operably-linked to a pro nucleotide sequence. These polynucleotide cassettes are useful for making multimeric proteins. Each of the genes of interest encodes a polypeptide that forms a part of the multimeric protein. These polynucleotide cassettes are administered to an individual for expression of the polypeptide sequences and expression and/or formation of the multimeric protein. Preferably, the individual is an animal from which the proteins or multimeric protein can be harvested. Preferred animals are egg-laying or milk-producing animals. In some embodiments, the polynucleotide cassettes are located within transposon-based vectors that allow for incorporation of the cassettes into the DNA of the individual.

The pro polynucleotide sequences operably-linked to the genes of interest include pro portions of prepro polynucleotide sequences commonly associated with polynucleotides encoding proteins secreted from a cell in nature. It may be that the pre polynucleotide sequence functions to direct the resultant protein into the endoplasmic reticulum and the pro sequence is cleaved within the endoplasmic reticulum or Golgi complex of a cell containing the protein. While prepro

polynucleotide sequences are associated with secreted polypeptides in nature, one discovery of the present invention is the use of pro portions of the prepro signal sequences to facilitate appropriate processing, expression, and/or formation of proteins and multimeric proteins, and more particularly, associated multimeric proteins. In the present invention, each gene of interest is operably-linked with a pro polynucleotide sequence. Figure 1 shows schematically one polynucleotide cassette containing two genes of interest, wherein each gene of interest is operably-linked to a pro polynucleotide sequence. The first gene of interest is operably-linked to a pro polynucleotide sequence that is part of a prepro polynucleotide sequence, while the second gene of interest is operably-linked to a pro polynucleotide sequence that is not part of a prepro polynucleotide sequence, but may have been derived from a prepro polynucleotide sequence. Accordingly, the term "pro sequence" encompasses a pro sequence that is part of a prepro sequence and a pro sequence that is not part of a prepro sequence, but may have been derived from a prepro sequence. In preferred embodiments, the most 5' pro polynucleotide sequence in the polynucleotide cassette is a part of a prepro polynucleotide sequence.

Several examples of prepro polynucleotides from which a pro polynucleotide can be derived or be a part of are a cecropin prepro, lysozyme prepro, ovomucin prepro, ovotransferrin prepro, a signal peptide for tumor necrosis factor receptor (SEQ ID NO:6), a signal peptide encoded by a polynucleotide sequence provided in one of SEQ ID NOs:7-54 and a signal peptide provide in SEQ ID NO:55. The prepro or pro polynucleotide can be a cecropin prepro or pro polynucleotide selected from the group consisting of cecropin A1, cecropin A2, cecropin B, cecropin C, cecropin D, cecropin E and cecropin F. In a preferred embodiment, the pro polynucleotide is a cecropin B pro polynucleotide having a sequence shown in SEQ ID NO:1 or SEQ ID NO:2. A preferred prepro polynucleotide is a cecropin B polynucleotide having a sequence shown in SEQ ID NO:3 or SEQ ID NO:4.

Figure 1 provides one embodiment of the invention wherein the polynucleotide cassette includes two genes of interest and two pro polynucleotide sequences arranged in the following order: a prepro polynucleotide, a first gene of interest, a pro polynucleotide, and a second gene of interest. Preferably, the sequences are arranged in the aforementioned order beginning at a 5' end of the polynucleotide cassette. Figure 2 provides a more specific embodiment of the present

invention wherein the first and second genes of interest are polynucleotides encoding antibody heavy and light chains. However, the invention includes polynucleotide cassettes containing at least two genes of interest. Each of the genes of interest is operably-linked to a pro polynucleotide. Each of these pro polynucleotides can be the same, or each can be different. In one embodiment, all of the pro polynucleotides in the polynucleotide cassette are the same and are cecropin pro polynucleotides. The most 5' cecropin pro polynucleotide is preferably a part of a cecropin prepro polynucleotide sequence as shown in Figure 3.

The polynucleotide cassettes of the present invention may be administered to an individual for production of a protein or a multimeric protein in that individual. The discussion that follows is intented to encompass methods for making one protein or a multimeric protein. Accordingly, the present invention includes a method of producing a protein in an individual comprising administering to the individual a polynucleotide cassette comprising one gene of interest encoding the protein, wherein the gene of interest is operably-linked to a pro polynucleotide sequence. The present invention includes a method of producing a multimeric protein in an individual comprising administering to the individual a polynucleotide cassette comprising at least two genes of interest, each encoding a part of the multimeric protein, wherein each gene of interest is operably-linked to a pro polynucleotide sequence. The present invention also includes a method of producing a multimeric protein in an individual comprising administering to the individual a polynucleotide cassette comprising a cecropin prepro sequence operably-linked to two or more genes of interest, each gene of interest encoding a part of the multimeric protein. This second method does not require the linking of pro polynucleotides to each gene of interest since the use of a cecropin prepro sequence itself in a polynucleotide cassette facilitates processing, expression, and/or formation of multimeric proteins. Polynucleotide cassettes containing the cecropin prepro polynucleotide can contain one or at least two genes of interest. Preferably, the cecropin prepro polynucleotide is located 5' of the one or more genes of interest in the polynucleotide cassette. One exemplary polynucleotide cassette is shown in Figure 3. In a preferred embodiment, the prepro sequence comprises a sequence shown in SEQ ID NO:3 or SEQ ID NO:4. As shown in Figure 3, the polynucleotide cassettes containing a cecropin prepro polynucleotide preferably contain a cleavage site between each of two genes of interest. Such cleavage site(s)

may be nucleotides encoding any cleavage sites including, but not limited to, an enzymatic cleavage site, a pro polynucleotide, and a photolabile cleavage site, a chemical cleavage site, and a self-splicing cleavage site (i.e., intein). Cleavage sites are discussed in more detail below.

The polynucleotide cassettes of the present invention are particularly suited for production of proteins and/or multimeric proteins in an individual. Individuals include both humans and animals. Preferred animals are egg-laying animals and milk-producing animals. As used herein, the term "egg-laying animal" includes all amniotes such as birds, turtles, lizards and monotremes. Monotremes are egg-laying mammals and include the platypus and echidna. The term "bird" or "fowl," as used herein, is defined as a member of the Aves class of animals which are characterized as warm-blooded, egg-laying vertebrates primarily adapted for flying. Avians include, without limitation, Ratites, Psittaciformes, Falconiformes, Piciformes, Strigiformes, Passeriformes, Coraciformes, Ralliformes, Cuculiformes, Columbiformes, Galliformes, Anseriformes, and Herodiones. The term "Ratite," as used herein, is defined as a group of flightless, mostly large, running birds comprising several orders and including the emus, ostriches, kiwis, and cassowaries. The term "Psittaciformes", as used herein, includes parrots and refers to a monofamilial order of birds that exhibit zygodactylism and have a strong hooked bill. A "parrot" is defined as any member of the avian family Psittacidae (the single family of the Psittaciformes), distinguished by the short, stout, strongly hooked beak. Preferred avians are poultry birds including chickens, quail, turkeys, geese and ducks. The term "chicken" as used herein denotes chickens used for table egg production, such as egg-type chickens, chickens reared for public meat consumption, or broilers, and chickens reared for both egg and meat production ("dual-purpose" chickens). The term "chicken" also denotes chickens produced by primary breeder companies, or chickens that are the parents, grandparents, great-grandparents, etc. of those chickens reared for public table egg, meat, or table egg and meat consumption.

When the polynucleotide cassettes of the present invention are administered to an egg-laying or milk-producing animal, a transgenic animal containing a polynucleotide cassette is created and the animal produces a transgenic protein or multimeric protein. It is preferred that the resultant protein or multimeric protein is deposited in the egg or in the milk. Various different signal sequences and promoters

may be used to achieve deposition of the protein or multimeric protein in the egg or in the milk and these are described in more detail below. In order to achieve a transgenic animal containing a polynucleotide cassette of the present invention, the polynucleotide cassettes can be administered to the individual with, or contained in, any vector, as naked DNA, or in any delivery construct or solution commonly known to one of ordinary skill in the art. A preferred vector for incorporation of the polynucleotide cassettes into an individual is a transposon-based vector described below.

Definitions

It is to be understood that as used in the specification and in the claims, "a" or "an" can mean one or more, depending upon the context in which it is used. Thus, for example, reference to "a cell" can mean that at least one cell can be utilized.

The term protein includes polypeptide and peptide and these terms are generally encompassed under the term "protein" and are also used interchangeablt herein. A multimeric protein includes more than one protein, polypeptide or peptide and is defined further below.

The term "antibody" is used interchangeably with the term "immunoglobulin" and is defined herein as a protein synthesized by an animal or a cell of the immune system in response to the presence of a foreign substance commonly referred to as an "antigen" or an "immunogen". The term antibody includes fragments of antibodies. Antibodies are characterized by specific affinity to a site on the antigen, wherein the site is referred to an "antigenic determinant" or an "epitope". Antigens can be naturally occurring or artificially engineered. Preferred antigens are those from organisms against which a vaccine is desired. Antigens and suspected or known epitopes may be obtained from any organism, including but not limited to those recited herein.

Artificially engineered antigens include but are not limited to small molecules, such as small peptides, attached to haptens such as macromolecules, for example proteins, nucleic acids, or polysaccharides. Artificially designed or engineered variants of naturally occurring antibodies and artificially designed or engineered antibodies not occurring in nature are all included in the current definition. Such variants include conservatively substituted amino acids and other forms of substitution as described in the section concerning proteins and polypeptides.

The term "egg" is defined herein as including a large female sex cell enclosed in a porous, calcarous or leathery shell, produced by birds and reptiles. The term "ovum" is defined as a female gamete, and is also known as an egg. Therefore, egg production in all animals other than birds and reptiles, as used herein, is defined as the production and discharge of an ovum from an ovary, or "ovulation". Accordingly, it is to be understood that the term "egg" as used herein is defined as a large female sex cell enclosed in a porous, calcarous or leathery shell, when a bird or reptile produces it, or it is an ovum when it is produced by all other animals.

The term "gene" is defined herein to include a polynucleotide that includes a coding region for a protein, peptide or polypeptide, with or without intervening sequences such as introns.

The term "multimeric protein" is defined herein to include one or more polypeptides that are associated, or joined, by any means including disulfde bonds. An example of this type of multimeric protein is an antibody that contains both heavy and light chains that are associated by disulfide bonds. These multimeric proteins are referred to herein as "associated multimeric proteins." The term "multimeric protein" also includes a polypeptide that is encoded by more than one gene of interest, such multimeric protein containg desired amino acid sequences encoded by the more than one gene of interest. An example of this type of multimeric protein is a single polypeptide containing a heavy chain polypeptide (first polypeptide of interest) and a light chain polypeptide (second polypeptide of interest). In these embodiments, the different polypeptides of interest may be separated by other polypeptide sequences such as spacer polypeptides and cleavage site polypeptides. These types of multimeric proteins are referred to herein as "multivalent multimeric proteins."

The term "milk-producing animal" refers herein to mammals including, but not limited to, bovine, ovine, porcine, equine, and primate animals. Milk-producing animals include but are not limited to cows, llamas, camels, goats, reindeer, zebu, water buffalo, yak, horses, pigs, rabbits, non-human primates, and humans.

The term "transgenic animal" refers to an animal having at least a portion of the transposon-based vector DNA is incorporated into its DNA. While a transgenic animal includes an animal wherein the transposon-based vector DNA is incorporated into the germline DNA, a transgenic animal also includes an animal having DNA in one or more somatic cells that contain a portion of the transposon-based vector DNA

for any period of time. In a preferred embodiment, a portion of the transposon-based vector comprises a gene of interest. More preferably, the gene of interest is incorporated into the animal's DNA for a period of at least five days, more preferably the laying life of the animal, and most preferably the life of the animal. In a further preferred embodiment, the animal is an avian.

The term "vector" is used interchangeably with the terms "construct", "DNA construct" and "genetic construct" to denote synthetic nucleotide sequences used for manipulation of genetic material, including but not limited to cloning, subcloning, sequencing, or introduction of exogenous genetic material into cells, tissues or organisms, such as birds. It is understood by one skilled in the art that vectors may contain synthetic DNA sequences, naturally occurring DNA sequences, or both. The vectors of the present invention are transposon-based vectors as described herein.

When referring to two nucleotide sequences, one being a regulatory sequence, the term "operably-linked" is defined herein to mean that the two sequences are associated in a manner that allows the regulatory sequence to affect expression of the other nucleotide sequence. It is not required that the operably-linked sequences be directly adjacent to one another with no intervening sequence(s).

The term "regulatory sequence" is defined herein as including promoters, enhancers and other expression control elements such as polyadenylation sequences, matrix attachment sites, insulator regions for expression of multiple genes on a single construct, ribosome entry/attachment sites, introns that are able to enhance expression, and silencers.

Transposon-Based Vectors

While not wanting to be bound by the following statement, it is believed that the nature of the DNA construct is an important factor in successfully producing transgenic animals. The "standard" types of plasmid and viral vectors that have previously been almost universally used for transgenic work in all species, especially avians, have low efficiencies and may constitute a major reason for the low rates of transformation previously observed. The DNA (or RNA) constructs previously used often do not integrate into the host DNA, or integrate only at low frequencies. Other factors may have also played a part, such as poor entry of the vector into target cells. The present invention provides transposon-based vectors that can be administered to an animal that overcome the prior art problems relating to low transgene integration

frequencies. Two preferred transposon-based vectors of the present invention in which a transposase, gene of interest and other polynucleotide sequences may be introduced are termed pTnMCS (SEQ ID NO:56) and pTnMod (SEQ ID NO:57).

The transposon-based vectors of the present invention produce integration frequencies an order of magnitude greater than has been achieved with previous vectors. More specifically, intratesticular injections performed with a prior art transposon-based vector (described in U.S. Patent No. 5,719,055) resulted in 41% sperm positive roosters whereas intratesticular injections performed with the novel transposon-based vectors of the present invention resulted in 77% sperm positive roosters. Actual frequencies of integration were estimated by either or both comparative strength of the PCR signal from the sperm and histological evaluation of the testes and sperm by quantitative PCR.

The transposon-based vectors of the present invention include a transposase gene operably-linked to a first promoter, and a coding sequence for a desired protein or peptide operably-linked to a second promoter, wherein the coding sequence for the desired protein or peptide and its operably-linked promoter are flanked by transposase insertion sequences recognized by the transposase. The transposon-based vector also includes one or more of the following characteristics: a) one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:5) at the 3' end of the first promoter to enhance expression of the transposase; b) modifications of one or more of the codons for the first several N-terminal amino acids of the transposase, wherein the third base of each codon was changed to an A or a T without changing the corresponding amino acid; c) addition of one or more stop codons to enhance the termination of transposase synthesis; and/or, d) addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of the transposase gene. In one embodiment, the transposon-based vector comprises an avian optimized polyA sequence and does not comprise a modified Kozak sequence comprising ACCATG (SEQ ID NO:5). One example of such a transposon-based vector is the pTnMCS vector (SEQ ID NO:56). In another embodiment the transposon-based vector comprises a) one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:5) at the 3' end of the first promoter to enhance expression of the transposase; b) modifications of the codons for the first several N-terminal amino acids of the transposase, wherein the third base of each codon was changed to an A or a T without

changing the corresponding amino acid; c) addition of one or more stop codons to enhance the termination of transposase synthesis; and, d) addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of the transposase gene. One example of such a transposon-based vector is the pTnMod vector (SEQ ID NO:57). The transposon-based vector may additionally or alternatively include one or more of the following Kozak sequences at the 3' end of any promoter, including the promoter operably-linked to the transposase: ACCATGG (SEQ ID NO:58), AAGATGT (SEQ ID NO:59), ACGATGA (SEQ ID NO:60), AAGATGG (SEQ ID NO:61), GACATGA (SEQ ID NO:62), ACCATGA (SEQ ID NO:63), and ACCATGA (SEQ ID NO:64), ACCATGT (SEQ ID NO:65).

Transposases and Insertion Sequences

In a further embodiment of the present invention, the transposase found in the transposase-based vector is an altered target site (ATS) transposase and the insertion sequences are those recognized by the ATS transposase. However, the transposase located in the transposase-based vectors is not limited to a modified ATS transposase and can be derived from any transposase. Transposases known in the prior art include those found in AC7, Tn5SEQ1, Tn916, Tn951, Tn1721, Tn 2410, Tn1681, Tn1, Tn2, Tn3, Tn4, Tn5, Tn6, Tn9, Tn10, Tn30, Tn101, Tn903, Tn501, Tn1000 (γδ), Tn1681, Tn2901, AC transposons, Mp transposons, Spm transposons, En transposons, Dotted transposons, Mu transposons, Ds transposons, dSpm transposons and I transposons. According to the present invention, these transposases and their regulatory sequences are modified for improved functioning as follows: a) the addition one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:5) at the 3' end of the promoter operably-linked to the transposase; b) a change of one or more of the codons for the first several amino acids of the transposase, wherein the third base of each codon was changed to an A or a T without changing the corresponding amino acid; c) the addition of one or more stop codons to enhance the termination of transposase synthesis; and/or, d) the addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of the transposase gene.

Although not wanting to be bound by the following statement, it is believed that the modifications of the first several N-terminal codons of the transposase gene facilitate transcription of the transposase gene, in part, by increasing strand dissociation during transcription. It is preferable that one or more of between

approximately the first 1 to 20, more preferably 3 to 15, and most preferably between 4 to 12 N-terminal codons of the transposase are modified such that the third base of each codon is changed to an A or a T without changing the encoded amino acid. In one embodiment, the first ten N-terminal codons of the transposase gene are modified in this manner. It is also preferred that the transposase contain mutations that make it less specific for preferred insertion sites and thus increases the rate of transgene insertion as discussed in U.S. Patent No. 5,719,055.

In some embodiments, the transposon-based vectors are optimized for expression in a particular host by changing the methylation patterns of the vector DNA. For example, prokaryotic methylation may be reduced by using a methylation deficient organism for production of the transposon-based vector. The transposon-based vectors may also be methylated to resemble eukaryotic DNA for expression in a eukaryotic host.

Transposases and insertion sequences from other analogous eukaryotic transposon-based vectors that can also be modified and used are, for example, the Drosophila P element derived vectors disclosed in U.S. Patent No. 6,291,243; the Drosophila mariner element described in Sherman et al. (1998); or the sleeping beauty transposon. See also Hackett et al. (1999); D. Lampe et al., 1999. Proc. Natl. Acad. Sci. USA, 96:11428-11433; S. Fischer et al., 2001. Proc. Natl. Acad. Sci. USA, 98:6759-6764; L. Zagoraiou et al., 2001. Proc. Natl. Acad. Sci. USA, 98:11474-11478; and D. Berg et al. (Eds.), Mobile DNA, Amer. Soc. Microbiol. (Washington, D.C., 1989). However, it should be noted that bacterial transposon-based elements are preferred, as there is less likelihood that a eukaryotic transposase in the recipient species will recognize prokaryotic insertion sequences bracketing the transgene.

Many transposases recognize different insertion sequences, and therefore, it is to be understood that a transposase-based vector will contain insertion sequences recognized by the particular transposase also found in the transposase-based vector. In a preferred embodiment of the invention, the insertion sequences have been shortened to about 70 base pairs in length as compared to those found in wild-type transposons that typically contain insertion sequences of well over 100 base pairs.

While the examples provided below incorporate a "cut and insert" Tn10 based vector that is destroyed following the insertion event, the present invention also encompasses the use of a "rolling replication" type transposon-based vector. Use of a

rolling replication type transposon allows multiple copies of the transposon/transgene to be made from a single transgene construct and the copies inserted. This type of transposon-based system thereby provides for insertion of multiple copies of a transgene into a single genome. A rolling replication type transposon-based vector may be preferred when the promoter operably-linked to gene of interest is endogenous to the host cell and present in a high copy number or highly expressed. However, use of a rolling replication system may require tight control to limit the insertion events to non-lethal levels. Tn1, Tn2, Tn3, Tn4, Tn5, Tn9, Tn21, Tn501, Tn551, Tn951, Tn1721, Tn2410 and Tn2603 are examples of a rolling replication type transposon, although Tn5 could be both a rolling replication and a cut and insert type transposon.

Stop Codons and PolyA Sequences

In one embodiment, the transposon-based vector contains two stop codons operably-linked to the transposase and/or to the gene of interest. In an alternate embodiment, one stop codon of UAA or UGA is operably linked to the transposase and/or to the gene of interest.

As used herein an "effective polyA sequence" refers to either a synthetic or non-synthetic sequence that contains multiple and sequential nucleotides containing an adenine base (an A polynucleotide string) and that increases expression of the gene to which it is operably-linked. A polyA sequence may be operably-linked to any gene in the transposon-based vector including, but not limited to, a transposase gene and a gene of interest. A preferred polyA sequence is optimized for use in the host animal or human. In one embodiment, the polyA sequence is optimized for use in an avian species and more specifically, a chicken. An avian optimized polyA sequence generally contains a minimum of 40 base pairs, preferably between approximately 40 and several hundred base pairs, and more preferably approximately 75 base pairs that precede the A polynucleotide string and thereby separate the stop codon from the A polynucleotide string. In one embodiment of the present invention, the polyA sequence comprises a conalbumin polyA sequence as provided in SEQ ID NO:66 and as taken from GenBank accession # Y00407, base pairs 10651-11058. In another embodiment, the polyA sequence comprises a synthetic polynucleotide sequence shown in SEQ ID NO:67. In yet another embodiment, the polyA sequence comprises an avian optimized polyA sequence provided in SEQ ID NO:68. A chicken optimized

polyA sequence may also have a reduced amount of CT repeats as compared to a synthetic polyA sequence.

It is a suprising discovery of the present invention that such an avian optimized poly A sequence increases expression of a polynucleotide to which it is operably-linked in an avian as compared to a non-avian optimized polyA sequence. Accordingly, the present invention includes methods of or increasing incorporation of a gene of interest wherein the gene of interest resides in a transposon-based vector containing a transposase gene and wherein the transposase gene is operably linked to an avian optimized polyA sequence. The present invention also includes methods of increasing expression of a gene of interest in an avian that includes administering a gene of interest to the avian, wherein the gene of interest is operably-linked to an avian optimized polyA sequence. An avian optimized polyA nucleotide string is defined herein as a polynucleotide containing an A polynucleotide string and a minimum of 40 base pairs, preferably between approximately 40 and several hundred base pairs, and more preferably approximately 75 base pairs that precede the A polynucleotide string. The present invention further provides transposon-based vectors containing a gene of interest or transposase gene operably linked to an avian optimized polyA sequence.

Promoters and Enhancers

The first promoter operably-linked to the transposase gene and the second promoter operably-linked to the gene of interest can be a constitutive promoter or an inducible promoter. Constitutive promoters include, but are not limited to, immediate early cytomegalovirus (CMV) promoter, herpes simplex virus 1 (HSV1) immediate early promoter, SV40 promoter, lysozyme promoter, early and late CMV promoters, early and late HSV promoters, β-actin promoter, tubulin promoter, Rous-Sarcoma virus (RSV) promoter, and heat-shock protein (HSP) promoter. Inducible promoters include tissue-specific promoters, developmentally-regulated promoters and chemically inducible promoters. Examples of tissue-specific promoters include the glucose 6 phosphate (G6P) promoter, vitellogenin promoter, ovalbumin promoter, ovomucoid promoter, conalbumin promoter, ovotransferrin promoter, prolactin promoter, kidney uromodulin promoter, and placental lactogen promoter. In one embodiment, the vitellogenin promoter includes a polynucleotide sequence of SEQ ID NO:69. The G6P promoter sequence may be deduced from a rat G6P gene

untranslated upstream region provided in GenBank accession number U57552.1. Examples of developmentally-regulated promoters include the homeobox promoters and several hormone induced promoters. Examples of chemically inducible promoters include reproductive hormone induced promoters and antibiotic inducible promoters such as the tetracycline inducible promoter and the zinc-inducible metallothionine promoter.

Other inducible promoter systems include the Lac operator repressor system inducible by IPTG (isopropyl beta-D-thiogalactoside) (Cronin, A. et al. 2001. Genes and Development, v. 15), ecdysone-based inducible systems (Hoppe, U. C. et al. 2000. Mol. Ther. 1:159-164); estrogen-based inducible systems (Braselmann, S. et al. 1993. Proc. Natl. Acad. Sci. 90:1657-1661); progesterone-based inducible systems using a chimeric regulator, GLVP, which is a hybrid protein consisting of the GAL4 binding domain and the herpes simplex virus transcriptional activation domain, VP16, and a truncated form of the human progesterone receptor that retains the ability to bind ligand and can be turned on by RU486 (Wang, et al. 1994. Proc. Natl. Acad. Sci. 91:8180-8184); CID-based inducible systems using chemical inducers of dimerization (CIDs) to regulate gene expression, such as a system wherein rapamycin induces dimerization of the cellular proteins FKBP12 and FRAP (Belshaw, P. J. et al. 1996. J. Chem. Biol. 3:731-738; Fan, L. et al. 1999. Hum. Gene Ther. 10:2273-2285; Shariat, S.F. et al. 2001. Cancer Res. 61:2562-2571; Spencer, D.M. 1996. Curr. Biol. 6:839-847). Chemical substances that activate the chemically inducible promoters can be administered to the animal containing the transgene of interest via any method known to those of skill in the art.

Other examples of cell or tissue-specific and constitutive promoters include but are not limited to smooth-muscle SM22 promoter, including chimeric SM22alpha/telokin promoters (Hoggatt A.M. et al., 2002. Circ Res. 91(12):1151-9); ubiquitin C promoter (Biochim Biophys Acta, 2003. Jan. 3;1625(1):52-63); Hsf2 promoter; murine COMP (cartilage oligomeric matrix protein) promoter; early B cell-specific mb-1 promoter (Sigvardsson M., et al., 2002. Mol. Cell Biol. 22(24):8539-51); prostate specific antigen (PSA) promoter (Yoshimura I. et al., 2002, J. Urol. 168(6):2659-64); exorh promoter and pineal expression-promoting element (Asaoka Y., et al., 2002. Proc. Natl. Acad. Sci. 99(24):15456-61); neural and liver ceramidase gene promoters (Okino N. et al., 2002. Biochem. Biophys. Res. Commun.

299(1):160-6); PSP94 gene promoter/enhancer (Gabril M.Y. et al., 2002. Gene Ther. 9(23):1589-99); promoter of the human FAT/CD36 gene (Kuriki C., et al., 2002. Biol. Pharm. Bull. 25(11):1476-8); VL30 promoter (Staplin W.R. et al., 2002. Blood October 24, 2002); and IL-10 promoter (Brenner S., et al., 2002. J. Biol. Chem. December 18, 2002).

Examples of avian promoters include, but are not limited to, promoters controlling expression of egg white proteins, such as ovalbumin, ovotransferrin (conalbumin), ovomucoid, lysozyme, ovomucin, g2 ovoglobulin, g3 ovoglobulin, ovoflavoprotein, ovostatin (ovomacroglobin), cystatin, avidin, thiamine-binding protein, glutamyl aminopeptidase minor glycoprotein 1, minor glycoprotein 2; and promoters controlling expression of egg-yolk proteins, such as vitellogenin, very low-density lipoproteins, low density lipoprotein, cobalamin-binding protein, riboflavin-binding protein, biotin-binding protein (Awade, 1996. Z. Lebensm. Unters. Forsch. 202:1-14). An advantage of using the vitellogenin promoter is that it is active during the egg-laying stage of an animal's life-cycle, which allows for the production of the protein of interest to be temporally connected to the import of the protein of interest into the egg yolk when the protein of interest is equipped with an appropriate targeting sequence. In some embodiments, the avian promoter is an oviduct-specific promoter. As used herein, the term "oviduct-specific promoter" includes, but is not limited to, ovalbumin, ovotransferrin (conalbumin), ovomucoid, lysozyme, ovomucin, g2 ovoglobulin, g3 ovoglobulin, ovoflavoprotein, and ovostatin (ovomacroglobin) promoters.

Liver-specific promoters of the present invention include, but are not limited to, the following promoters, vitellogenin promoter, G6P promoter, cholesterol-7-alpha-hydroxylase (CYP7A) promoter, phenylalanine hydroxylase (PAH) promoter, protein C gene promoter, insulin-like growth factor I (IGF-I) promoter, bilirubin UDP-glucuronosyltransferase promoter, aldolase B promoter, furin promoter, metallothioneine promoter, albumin promoter, and insulin promoter.

Also included in the present invention are promoters that can be used to target expression of a protein of interest into the milk of a milk-producing animal including, but not limited to, β lactoglobin promoter, whey acidic protein promoter, lactalbumin promoter and casein promoter.

Promoters associated with cells of the immune system may also be used. Acute phase promoters such as interleukin (IL)-1 and IL-2 may be employed. Promoters for heavy and light chain Ig may also be employed. The promoters of the T cell receptor components CD4 and CD8, B cell promoters and the promoters of CR2 (complement receptor type 2) may also be employed. Immune system promoters are preferably used when the desired protein is an antibody protein.

Also included in this invention are modified promoters/enhancers wherein elements of a single promoter are duplicated, modified, or otherwise changed. In one embodiment, steroid hormone-binding domains of the ovalbumin promoter are moved from about -6.5 kb to within approximately the first 1000 base pairs of the gene of interest. Modifying an existing promoter with promoter/enhancer elements not found naturally in the promoter, as well as building an entirely synthetic promoter, or drawing promoter/enhancer elements from various genes together on a non-natural backbone, are all encompassed by the current invention.

Accordingly, it is to be understood that the promoters contained within the transposon-based vectors of the present invention may be entire promoter sequences or fragments of promoter sequences. For example, in one embodiment, the promoter operably linked to a gene of interest is an approximately 900 base pair fragment of a chicken ovalbumin promoter (SEQ ID NO:70). The constitutive and inducible promoters contained within the transposon-based vectors may also be modified by the addition of one or more modified Kozak sequences of ACCATG (SEQ ID NO:5).

As indicated above, the present invention includes transposon-based vectors containing one or more enhancers. These enhancers may or may not be operably-linked to their native promoter and may be located at any distance from their operably-linked promoter. A promoter operably-linked to an enhancer is referred to herein as an "enhanced promoter." The enhancers contained within the transposon-based vectors are preferably enhancers found in birds, and more preferably, an ovalbumin enhancer, but are not limited to these types of enhancers. In one embodiment, an approximately 675 base pair enhancer element of an ovalbumin promoter is cloned upstream of an ovalbumin promoter with 300 base pairs of spacer DNA separating the enhancer and promoter. In one embodiment, the enhancer used as a part of the present invention comprises base pairs 1-675 of a Chicken Ovalbumin

enhancer from GenBank accession #S82527.1. The polynucleotide sequence of this enhancer is provided in SEQ ID NO:71.

Also included in some of the transposon-based vectors of the present invention are cap sites and fragments of cap sites. In one embodiment, approximately 50 base pairs of a 5' untranslated region wherein the capsite resides are added on the 3' end of an enhanced promoter or promoter. An exemplary 5' untranslated region is provided in SEQ ID NO:72. A putative cap-site residing in this 5' untranslated region preferably comprises the polynucleotide sequence provided in SEQ ID NO:73.

In one embodiment of the present invention, the first promoter operably-linked to the transposase gene is a constitutive promoter and the second promoter operably-linked to the gene of interest is a tissue-specific promoter. In the second embodiment, use of the first constitutive promoter allows for constitutive activation of the transposase gene and incorporation of the gene of interest into virtually all cell types, including the germline of the recipient animal. Although the gene of interest is incorporated into the germline generally, the gene of interest is only expressed in a tissue-specific manner. A transposon-based vector having a constitutive promoter operably-linked to the transposase gene can be administered by any route, and in one embodiment, the vector is administered to an ovary or to an artery leading to the ovary. In another embodiment, the vector is administered into the lumen of the oviduct or into an artery supplying the oviduct.

It should be noted that cell- or tissue-specific expression as described herein does not require a complete absence of expression in cells or tissues other than the preferred cell or tissue. Instead, "cell-specific" or "tissue-specific" expression refers to a majority of the expression of a particular gene of interest in the preferred cell or tissue, respectively.

When incorporation of the gene of interest into the germline is not preferred, the first promoter operably-linked to the transposase gene can be a tissue-specific promoter. For example, transfection of a transposon-based vector containing a transposase gene operably-linked to an oviduct specific promoter such as the ovalbumin promoter provides for activation of the transposase gene and incorporation of the gene of interest in the cells of the oviduct but not into the germline and other cells generally. In this embodiment, the second promoter operably-linked to the gene of interest can be a constitutive promoter or an inducible promoter. In a preferred

embodiment, both the first promoter and the second promoter are an ovalbumin promoter. In embodiments wherein tissue-specific expression or incorporation is desired, it is preferred that the transposon-based vector is administered directly to the tissue of interest or to an artery leading to the tissue of interest. In a preferred embodiment, the tissue of interest is the oviduct and administration is achieved by direct injection into the lumen of the oviduct or an artery leading to the oviduct. In a further preferred embodiment, administration is achieved by direct injection into the lumen of the magnum or the infundibulum of the oviduct.

Accordingly, cell specific promoters may be used to enhance transcription in selected tissues. In birds, for example, promoters that are found in cells of the fallopian tube, such as ovalbumin, conalbumin, ovomucoid and/or lysozyme, are used in the vectors to ensure transcription of the gene of interest in the epithelial cells and tubular gland cells of the fallopian tube, leading to synthesis of the desired protein encoded by the gene and deposition into the egg white. In mammals, promoters specific for the epithelial cells of the alveoli of the mammary gland, such as prolactin, insulin, beta lactoglobin, whey acidic protein, lactalbumin, casein, and/or placental lactogen, are used in the design of vectors used for transfection of these cells for the production of desired proteins for deposition into the milk. In liver cells, the G6P promoter may be employed to drive transcription of the gene of interest for protein production. Proteins made in the liver of birds may be delivered to the egg yolk.

In order to achieve higher or more efficient expression of the transposase gene, the promoter and other regulatory sequences operably-linked to the transposase gene may be those derived from the host. These host specific regulatory sequences can be tissue specific as described above or can be of a constitutive nature. For example, an avian actin promoter and its associated polyA sequence can be operably-linked to a transposase in a transposase-based vector for transfection into an avian. Examples of other host specific promoters that could be operably-linked to the transposase include the myosin and DNA or RNA polymerase promoters.

Directing Sequences

In some embodiments of the present invention, the gene of interest is operably-linked to a directing sequence or a sequence that provides proper conformation to the desired protein encoded by the gene of interest. As used herein, the term "directing sequence" refers to both signal sequences and targeting sequences.

An egg directing sequence includes, but is not limited to, an ovomucoid signal sequence, an ovalbumin signal sequence, a cecropin pre pro signal sequence, and a vitellogenin targeting sequence. The term "signal sequence" refers to an amino acid sequence, or the polynucleotide sequence that encodes the amino acid sequence, a portion or the entirety of which directs the protein to which it is linked to the endoplasmic reticulum in a eukaryote, and more preferably the translocational pores in the endoplasmic reticulum, or the plasma membrane in a prokaryote, or mitochondria, such as for the purpose of gene therapy for mitochondrial diseases. Signal and targeting sequences can be used to direct a desired protein into, for example, the milk, when the transposon-based vectors are administered to a milk-producing animal.

Signal sequences can also be used to direct a desired protein into, for example, a secretory pathway for incorporation into the egg yolk or the egg white, when the transposon-based vectors are administered to a bird or other egg-laying animal. The present invention also includes a gene of interest operably-linked to a second gene containing a signal sequence. An example of such an embodiment is wherein the gene of interest is operably-linked to the ovalbumin gene that contains an ovalbumin signal sequence. Other signal sequences that can be included in the transposon-based vectors include, but are not limited to the ovotransferrin and lysozyme signal sequences. In one embodiment, the signal sequence is an ovalbumin signal sequence including a sequence shown in SEQ ID NO:74. In another embodiment, the signal sequence is a shortened ovalbumin signal sequence including a sequence shown in SEQ ID NO:75 or SEQ ID NO:76.

As also used herein, the term "targeting sequence" refers to an amino acid sequence, or the polynucleotide sequence encoding the amino acid sequence, which amino acid sequence is recognized by a receptor located on the exterior of a cell. Binding of the receptor to the targeting sequence results in uptake of the protein or peptide operably-linked to the targeting sequence by the cell. One example of a targeting sequence is a vitellogenin targeting sequence that is recognized by a vitellogenin receptor (or the low density lipoprotein receptor) on the exterior of an oocyte. In one embodiment, the vitellogenin targeting sequence includes the polynucleotide sequence of SEQ ID NO:77. In another embodiment, the vitellogenin targeting sequence includes all or part of the vitellogenin gene. Other targeting

sequences include VLDL and Apo E, which are also capable of binding the vitellogenin receptor. Since the ApoE protein is not endogenously expressed in birds, its presence may be used advantageously to identify birds carrying the transposon-based vectors of the present invention.

Genes of Interest

The genes of interest in the polynucleotide cassette can be any gene, and preferably are genes that encode proteins or portions of multimeric proteins that may be used in vaccines. Such genes may be coding regions for desired epitopes in infectious organisms, including but not limited to viral, bacterial, protozoal, fungal, and parasitic organisms. Such genes are known to one of ordinary skill in the art and may be obtained from GenBank and from scientific and patent literature. A gene of interest may contain modifications of the codons for the first several N-terminal amino acids of the gene of interest, wherein the third base of each codon is changed to an A or a T without changing the corresponding amino acid.

In one embodiment, the genes of interest are antibody genes or portions of antibody genes. Figure 2 shows a schematic drawing of a polynucleotide cassette containing an antibody heavy chain and an antibody light chain as two genes of interest. Antibodies used in or encoded by the polynucleotide cassettes of the present invention include, but are not limited to, IgG, IgM, IgA, IgD, IgE, IgY, lambda chains, kappa chains, bi-specific antibodies, and fragments thereof; scFv fragments, Fc fragments, and Fab fragments as well as dimeric, trimeric and oligomeric forms of antibody fragments. Desired antibodies include, but are not limited to, naturally occurring antibodies, human antibodies, humanized antibodies, autoantibodies and hybrid antibodies. Preferred antibodies include those useful as vaccines against selected diseases. Genes encoding modified versions of naturally occurring antibodies or fragments thereof and genes encoding artificially designed antibodies or fragments thereof may be incorporated into the transposon-based vectors of the present invention. Desired antibodies also include antibodies with the ability to bind specific ligands, for example, antibodies against proteins associated with diseases, particularly infectious diseases. Accordingly, the present invention encompasses a polynucleotide cassette as described herein containing one or more genes encoding a heavy immunoglobulin (Ig) chain and a light Ig chain.

Antibodies that may be produced using the present invention include, but are not limited to, antibodies for providing passive immunity to an animal or a human against an infectious disease or a toxic agent. The antibodies prepared using the methods of the present invention may also be designed to possess specific labels that may be detected through means known to one of ordinary skill in the art. For example, antibodies may be labeled with a fluorescent label attached that may be detected following exposure to specific wavelengths. Such labeled antibodies may be primary antibodies directed to a specific antigen, for example, rhodamine-labeled rabbit anti-growth hormone, or may be labeled secondary antibodies, such as fluorescein-labeled goat-anti chicken IgG. Such labeled antibodies are known to one of ordinary skill in the art. The antibodies may also be designed to possess specific sequences useful for purification through means known to one of ordinary skill in the art. Labels useful for attachment to antibodies are also known to one of ordinary skill in the art. Some of these labels are described in the "Handbook of Fluorescent Probes and Research Products", ninth edition, Richard P. Haugland (ed) Molecular Probes, Inc. Eugene, OR), which is incorporated herein in its entirety. Proteins and multimeric proteins made according to the present invention may also be labeled with these methods. Antibodies, proteins and multimeric proteins produced with the present invention may be used as laboratory reagents for numerous applications including radioimmunoassay, western blots, dot blots, ELISA, immunoaffinity columns and other procedures requiring antibodies as known to one of ordinary skill in the art. Such antibodies include primary antibodies, secondary antibodies and tertiary antibodies, which may be labeled or unlabeled.

Additional antibodies that may be made with the practice of the present invention include, but are not limited to, primary antibodies, secondary antibodies, designer antibodies, anti-protein antibodies, anti-peptide antibodies, anti-DNA antibodies, anti-RNA antibodies, antibodies against non-natural antigens, anti-venom antibodies, antibodies directed against epitopes associated with infectious disease, including, anti-viral, anti-bacterial, anti-protozoal, anti-fungal, anti-parasitic, anti-receptor, anti-lipid, anti-phospholipid, anti-growth factor, anti-cytokine, anti-monokine, anti-idiotype, and anti-accessory (presentation) protein antibodies. Antibodies made with the present invention, as well as light chains or heavy chains, may also be used to inhibit enzyme activity.

Antibodies that may be produced using the present invention include, but are not limited to, antibodies made against the following proteins: Bovine γ-Globulin, Serum; Bovine IgG, Plasma; Chicken γ-Globulin, Serum; Human γ-Globulin, Serum; Human IgA, Plasma; Human IgA_1, Myeloma; Human IgA_2, Myeloma; Human IgA_2, Plasma; Human IgD, Plasma; Human IgE, Myeloma; Human IgG, Plasma; Human IgG, Fab Fragment, Plasma; Human IgG, $F(ab')_2$ Fragment, Plasma; Human IgG, Fc Fragment, Plasma; Human IgG_1, Myeloma; Human IgG_2, Myeloma; Human IgG_3, Myeloma; Human IgG_4, Myeloma; Human IgM, Myeloma; Human IgM, Plasma; Human Immunoglobulin, Light Chain κ, Urine; Human Immunoglobulin, Light Chains κ and λ, Plasma; Mouse γ-Globulin, Serum; Mouse IgG, Serum; Mouse IgM, Myeloma; Rabbit γ-Globulin, Serum; Rabbit IgG, Plasma; and Rat γ-Globulin, Serum. In one embodiment, the transposon-based vector comprises the coding sequence of light and heavy chains of a murine monoclonal antibody that shows specificity for human seminoprotein (GenBank Accession numbers AY129006 and AY129304 for the light and heavy chains, respectively).

A further non-limiting list of antibodies that recognize other antibodies and that may be produced using the present invention is as follows: Anti-Chicken IgG, heavy (H) & light (L) Chain Specific (Sheep); Anti-Goat γ-Globulin (Donkey); Anti-Goat IgG, Fc Fragment Specific (Rabbit); Anti-Guinea Pig γ-Globulin (Goat); Anti-Human Ig, Light Chain, Type κ Specific; Anti-Human Ig, Light Chain, Type λ Specific; Anti-Human IgA, α-Chain Specific (Goat); Anti-Human IgA, Fab Fragment Specific; Anti-Human IgA, Fc Fragment Specific; Anti-Human IgA, Secretory; Anti-Human IgE, ε-Chain Specific (Goat); Anti-Human IgE, Fc Fragment Specific; Anti-Human IgG, Fc Fragment Specific (Goat); Anti-Human IgG, γ-Chain Specific (Goat); Anti-Human IgG, Fc Fragment Specific; Anti-Human IgG, Fd Fragment Specific; Anti-Human IgG, H & L Chain Specific (Goat); Anti-Human IgG_1, Fc Fragment Specific; Anti-Human IgG_2, Fc Fragment Specific; Anti-Human IgG_2, Fd Fragment Specific; Anti-Human IgG_3, Hinge Specific; Anti-Human IgG_4, Fc Fragment Specific; Anti-Human IgM, Fc Fragment Specific; Anti-Human IgM, μ-Chain Specific; Anti-Mouse IgE, ε-Chain Specific; Anti-Mouse γ-Globulin (Goat); Anti-Mouse IgG, γ-Chain Specific (Goat); Anti-Mouse IgG, γ-Chain Specific (Goat) $F(ab')_2$ Fragment; Anti-Mouse IgG, H & L Chain Specific (Goat); Anti-Mouse IgM, μ-Chain Specific (Goat); Anti-Mouse IgM, H & L Chain Specific (Goat); Anti-Rabbit γ-Globulin

(Goat); Anti-Rabbit IgG, Fc Fragment Specific (Goat); Anti-Rabbit IgG, H & L Chain Specific (Goat); Anti-Rat γ-Globulin (Goat); Anti-Rat IgG, H & L Chain Specific; Anti-Rhesus Monkey γ-Globulin (Goat); and, Anti-Sheep IgG, H & L Chain Specific.

Antibodies may be made against peptides such as p146 (SEQ ID NO:78 amino acid sequence, SEQ ID NO:79, nucleotide sequence), and other lytic peptides or peptides associated with disease, particularly infectious disease.

The following is yet another non-limiting of antibodies that can be produced by the methods of present invention: abciximab (ReoPro), abciximab anti-platelet aggregation monoclonal antibody, anti-CD11a (hu1124), anti-CD18 antibody, anti-CD20 antibody, anti-cytomegalovirus (CMV) antibody, anti-digoxin antibody, anti-hepatitis B antibody, anti-HER-2 antibody, anti-idiotype antibody to GD3 glycolipid, anti-IgE antibody, anti-IL-2R antibody, antimetastatic cancer antibody (mAb 17-1A), anti-rabies antibody, anti-respiratory syncytial virus (RSV) antibody, anti-Rh antibody, anti-TCR, anti-TNF antibody, anti-VEGF antibody and Fab fragment thereof, rattlesnake venom antibody, black widow spider venom antibody, coral snake venom antibody, antibody against very late antigen-4 (VLA-4), C225 humanized antibody to EGF receptor, chimeric (human & mouse) antibody against TNFα, antibody directed against GPIIb/IIIa receptor on human platelets, gamma globulin, anti-hepatitis B immunoglobulin, human anti-D immunoglobulin, human antibodies against *S aureus*, human tetanus immunoglobulin, humanized antibody against the epidermal growth receptor-2, humanized antibody against the α subunit of the interleukin-2 receptor, humanized antibody CTLA4IG, humanized antibody to the IL-2 R α-chain, humanized anti-CD40-ligand monoclonal antibody (5c8), humanized mAb against the epidermal growth receptor-2, humanized mAb to rous sarcoma virus, humanized recombinant antibody (IgG1k) against respiratory syncytial virus (RSV), lymphocyte immunoglobulin (anti-thymocyte antibody), lymphocyte immunoglobulin, mAb against factor VII, MDX-210 bi-specific antibody against HER-2, MDX-22, MDX-220 bi-specific antibody against TAG-72 on tumors, MDX-33 antibody to FcγR1 receptor, MDX-447 bi-specific antibody against EGF receptor, MDX-447 bispecific humanized antibody to EGF receptor, MDX-RA immunotoxin (ricin A linked) antibody, Medi-507 antibody (humanized form of BTI-322) against CD2 receptor on T-cells, monoclonal antibody LDP-02, muromonab-CD3(OKT3)

antibody, OKT3 ("muromomab-CD3") antibody, PRO 542 antibody, ReoPro ("abciximab") antibody, and TNF-IgG fusion protein.

Antibodies may also be made, in accordance with the present invention, against the following non-limiting list of organisms associated with disease and against peptides and/or proteins or other epitopes from these organisms: Anthrax (*Bacillus anthracis*), *Bacillus anthracis* (anthrax), Botulism (*Clostridium botulinum* toxin), *Brucella* species (brucellosis), Brucellosis (*Brucella* species), *Burkholderia mallei* (glanders), *Burkholderia pseudomallei* (melioidosis), *Chlamydia psittaci* (psittacosis), Cholera (*Vibrio cholerae*), *Clostridium botulinum* toxin (botulism), *Clostridium perfringens* (Epsilon toxin), *Coxiella burnetii* (Q fever), *E. coli* O157:H7 (*Escherichia coli*), Emerging infectious diseases such as Nipah virus and hantavirus, Epsilon toxin of *Clostridium perfringens*, *Escherichia coli* O157:H7 (*E. coli*), Food safety threats (e.g., *Salmonella* species, *Escherichia coli* O157:H7, *Shigella*), *Francisella tularensis* (tularemia), Glanders (*Burkholderia mallei*), Melioidosis (*Burkholderia pseudomallei*), Plague (*Yersinia pestis*), Psittacosis (*Chlamydia psittaci*), Q fever (*Coxiella burnetii*), Ricin toxin from *Ricinus communis* (castor beans), *Rickettsia prowazekii* (typhus fever), *Salmonella* species (salmonellosis), *Salmonella* Typhi (typhoid fever), Salmonellosis (*Salmonella* species), *Shigella* (shigellosis), Shigellosis (*Shigella*), Smallpox (variola major), Staphylococcal enterotoxin B, Tularemia (*Francisella tularensis*), Typhoid fever (*Salmonella* Typhi), Typhus fever (*Rickettsia prowazekii*), Variola major (smallpox), *Vibrio cholerae* (cholera), Viral encephalitis (alphaviruses [e.g., Venezuelan equine encephalitis, eastern equine encephalitis, western equine encephalitis]), Viral hemorrhagic fevers (filoviruses [e.g., Ebola, Marburg] and arenaviruses [e.g., Lassa, Machupo]), Water safety threats (e.g., *Vibrio cholerae, Cryptosporidium parvum*), and *Yersinia pestis* (plague).

Another non-limiting list of the antibodies that may be produced using the present invention is provided in product catalogs of companies such as Phoenix Pharmaceuticals, Inc. (www.phoenixpeptide.com; 530 Harbor Boulevard, Belmont, CA), Peninsula Labs San Carlos CA, SIGMA, St.Louis, MO www.sigma-aldrich.com, Cappel ICN, Irvine, California, www.icnbiomed.com, and Calbiochem, La Jolla, California, www.calbiochem.com, which are all incorporated herein by reference in their entirety. The polynucleotide sequences encoding these antibodies

may be obtained from the scientific literature, from patents, and from databases such as GenBank. Alternatively, one of ordinary skill in the art may design the antibody polynucleotide sequence by choosing the codons that encode for each amino acid in the desired antibody.

As an example of protein production using the method of the present invention which may be applied to the production of proteins useful in vaccines, the gene of interest is a proinsulin gene and the desired molecule is insulin. Proinsulin consists of three parts: a C-peptide and two strands of amino acids (the alpha and beta chains) that later become linked together to form the insulin molecule. In these embodiments, proinsulin is expressed in the oviduct tubular gland cells and then deposited in the egg white. One example of a proinsulin polynucleotide sequence is shown in SEQ ID NO:80, wherein the C-peptide cleavage site spans from Arg at position 31 to Arg at position 65.

Further included in the present invention are genes of interet that encode proteins and peptides synthesized by the immune system including those synthesized by the thymus, lymph nodes, spleen, and the gastrointestinal associated lymph tissues (GALT) system. The immune system proteins and peptides proteins that can be made in transgenic animals using the polynucleotide cassettes of the present invention include, but are not limited to, alpha-interferon, beta-interferon, gamma-interferon, alpha-interferon A, alpha-interferon 1, G-CSF, GM-CSF, interlukin-1 (IL-1), IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, TNF-α, and TNF-β. Other cytokines included in the present invention include cardiotrophin, stromal cell derived factor, macrophage derived chemokine (MDC), melanoma growth stimulatory activity (MGSA), macrophage inflammatory proteins 1 alpha (MIP-1 alpha), 2, 3 alpha, 3 beta, 4 and 5.

Genes encoding lytic peptides such as p146 are also included in the genes of interest of the present invention. In one embodiment, the p146 peptide comprises an amino acid sequence of SEQ ID NO:78. The present invention also encompasses a polynucleotide cassette comprising a p146 nucleic acid having a sequence of SEQ ID NO:79.

Enzymes are another class of proteins that may be encoded by the polynucleotide cassettes of the present invention. Such enzymes include but are not limited to adenosine deaminase, alpha-galactosidase, cellulase, collagenase, dnaseI,

hyaluronidase, lactase, L-asparaginase, pancreatin, papain, streptokinase B, subtilisin, superoxide dismutase, thrombin, trypsin, urokinase, fibrinolysin, glucocerebrosidase and plasminogen activator. In some embodiments wherein the enzyme could have deleterious effects, additional amino acids and a protease cleavage site are added to the carboxy end of the enzyme of interest in order to prevent expression of a functional enzyme. Subsequent digestion of the enzyme with a protease results in activation of the enzyme. Preferred are enzymes associate with disease processes and the function of infectious organisms.

A non-limiting list of the peptides and proteins that may be encoded by the polynucleotide cassettes of the present invention is provided in product catalogs of companies such as Phoenix Pharmaceuticals, Inc. (www.phoenixpeptide.com; 530 Harbor Boulevard, Belmont, CA), Peninsula Labs (San Carlos CA), SIGMA, (St.Louis, MO www.sigma-aldrich.com), Cappel ICN (Irvine, California, www.icnbiomed.com), and Calbiochem (La Jolla, California, www.calbiochem.com). The polynucleotide sequences encoding these proteins and peptides of interest may be obtained from the scientific literature, from patents, and from databases such as GenBank. Alternatively, one of ordinary skill in the art may design the polynucleotide sequence to be incorporated into the genome by choosing the codons that encode for each amino acid in the desired protein or peptide.

Other desired proteins that may be encoded by the polynucleotide cassettes of the present invention include bacitracin, polymixin b, vancomycin, cyclosporine, anti-RSV antibody, alpha-1 antitrypsin (AAT), anti-cytomegalovirus antibody, anti-hepatitis antibody, anti-inhibitor coagulant complex, anti-rabies antibody, anti-Rh(D) antibody, adenosine deaminase, anti-digoxin antibody, antivenin crotalidae (rattlesnake venom antibody), antivenin latrodectus (black widow spider venom antibody), antivenin micrurus (coral snake venom antibody), aprotinin, corticotropin (ACTH), diphtheria antitoxin, lymphocyte immune globulin (anti-thymocyte antibody), protamine, thyrotropin, capreomycin, α-galactosidase, gramicidin, streptokinase, tetanus toxoid, tyrothricin, IGF-1, proteins of varicella vaccine, anti-TNF antibody, anti-IL-2r antibody, anti-HER-2 antibody, OKT3 ("muromonab-CD3") antibody, TNF-IgG fusion protein, ReoPro ("abciximab") antibody, ACTH fragment 1-24, desmopressin, gonadotropin-releasing hormone, histrelin, leuprolide, lypressin, nafarelin, peptide that binds GPIIb/GPIIIa on platelets (integrilin),

goserelin, capreomycin, colistin, anti-respiratory syncytial virus, lymphocyte immune globulin (Thymoglovin, Atgam), panorex, alpha-antitrypsin, botulinin, lung surfactant protein, tumor necrosis receptor-IgG fusion protein (enbrel), gonadorelin, proteins of influenza vaccine, proteins of rotavirus vaccine, proteins of haemophilus b conjugate vaccine, proteins of poliovirus vaccine, proteins of pneumococcal conjugate vaccine, proteins of meningococcal C vaccine, proteins of influenza vaccine, megakaryocyte growth and development factor (MGDF), neuroimmunophilin ligand-A (NIL-A), brain-derived neurotrophic factor (BDNF), glial cell line-derived neurotrophic factor (GDNF), leptin (native), leptin B, leptin C, IL-1RA (interleukin-1RA), R-568, novel erythropoiesis-stimulating protein (NESP), humanized mAb to rous sarcoma virus (MEDI-493), glutamyl-tryptophan dipeptide IM862, LFA-3TIP immunosuppressive, humanized anti-CD40-ligand monoclonal antibody (5c8), gelsonin enzyme, tissue factor pathway inhibitor (TFPI), proteins of meningitis B vaccine, antimetastatic cancer antibody (mAb 17-1A), chimeric (human & mouse) mAb against TNFα, mAb against factor VII, relaxin, capreomycin, glycopeptide (LY333328), recombinant human activated protein C (rhAPC), humanized mAb against the epidermal growth receptor-2, altepase, anti-CD20 antigen, C2B8 antibody, insulin-like growth factor-1, atrial natriuretic peptide (anaritide), tenectaplase, anti-CD11a antibody (hu 1124), anti-CD18 antibody, mAb LDP-02, anti-VEGF antibody, fab fragment of anti-VEGF Ab, APO2 ligand (tumor necrosis factor-related apoptosis-inducing ligand), rTGF-β (transforming growth factor-β), alpha-antitrypsin, ananain (a pineapple enzyme), humanized mAb CTLA4IG, PRO 542 (mAb), D2E7 (mAb), calf intestine alkaline phosphatase, α-L-iduronidase, α-L-galactosidase (humanglutamic acid decarboxylase, acid sphingomyelinase, bone morphogenetic protein-2 (rhBMP-2), proteins of HIV vaccine, T cell receptor (TCR) peptide vaccine, TCR peptides, V beta 3 and V beta 13.1. (IR502), (IR501), BI 1050/1272 mAb against very late antigen-4 (VLA-4), C225 humanized mAb to EGF receptor, anti-idiotype antibody to GD3 glycolipid, antibacterial peptide against *H. pylori*, MDX-447 bispecific humanized mAb to EGF receptor, anti-cytomegalovirus (CMV), Medi-491 B19 parvovirus vaccine, humanized recombinant mAb (IgG1k) against respiratory syncytial virus (RSV), urinary tract infection vaccine (against "pili" on *Escherechia coli* strains), proteins of lyme disease vaccine against *B. burgdorferi* protein (DbpA), proteins of Medi-501 human papilloma virus-11 vaccine (HPV), *Streptococcus pneumoniae* vaccine, Medi-507

mAb (humanized form of BTI-322) against CD2 receptor on T-cells, MDX-33 mAb to FcγR1 receptor, MDX-RA immunotoxin (ricin A linked) mAb, MDX-210 bi-specific mAb against HER-2, MDX-447 bi-specific mAb against EGF receptor, MDX-22, MDX-220 bi-specific mAb against TAG-72 on tumors, colony-stimulating factor (CSF) (molgramostim), humanized mAb to the IL-2 R α-chain (basiliximab), mAb to IgE (IGE 025A), myelin basic protein-altered peptide (MSP771A), humanized mAb against the epidermal growth receptor-2, humanized mAb against the α subunit of the interleukin-2 receptor, low molecular weight heparin, anti-hemophillic factor, and bactericidal/permeability-increasing protein (r-BPI).

The following paragraph describes specific examples of epitopes and antibodies associated with specific diseases that may be made as the proteins, multimeric proteins of the present invention. Such epitopes may also be used in antibody production. Anthrax: PA (protective antigen) mediates entry of toxin's lethal factor (LF) into cell, therefore PA is binding domain. There are at least three different binding sites on PA where LF can bind. Site I epitope IG3PA63, site II 2D3PA, 2D5PA, 10D2PA, and site III 3B6PA, 14B7PA, 10E10PA63. Botulism: BoNT (botulism neurotoxin), binding domains within neurotoxin A (BoNT/A) derived from segments of heavy chain and encoded proteins : H455-661, H1150-1289. Accession number X73844. BoNT/A in infants: Accession number X73423.

Tularemia: mAbs against LPS of francisella tularensis: FT14, FT2F11. OMP specific mAbs react to IgGl subclass of *F. tularensis*. Yersenia pestis: causes pneumonic and bubonic plague. LCrV: accession number M26405 (*Y. pestis*), LCrV: accession number M57893 (*Y. pseudotuberculosis*). Caf-1 encodes capsular antigen fraction 1 (F1) accession number X61996. Shigella: epitopes 2C32E6, 4D64B9 (reactive to *S. flexneri* and *S. boydii*). 5E45D8 – *S. flexneri*. 4B33D10, 1B52F10 – all species of Shieglla. Shigella monoclonal antibodies 2C32E6, 4D64B9, 5E45D8 (reactive with S. Felxneri). 4B33D10 and 1B52F1 (species of Shigella) and 1B52F10 all strains (see Rahman & Stimson, Hybridoma, 20:85-90, 2001). See Venkatesan et al., PNAS 85:9317-21 (1988) accession number J04117 S. flexnari). IpaD between amino acid residues 14 and 44 of Shigella flexnari. IpaD is against the unique region of this protein exposed on the surface of the pathogen (Turbyfill et al., Infect. Immunity 66:1999-2006 (1998). See Muller-Loennies et al., J. Biol. Chem. 278: 25618-27 (2003) for epitopes related to Shigella. LcrV, Yersinia pestis (accession.

number M26405), Yersinia pseudotuberculosis (accessionnumber M57893, Yersinia enterocolitica. Fusion of IcrV and the structural gene for protein A. One protective epitope resides internally between amino acids 168 and 275 (Motin et al., Infect. Immunity 62:41920-201 (1994). For additional epitopes on Yersinia pastis see accession number M26405 LcrV (Price et al., J. Bacteriol. 171: 5646-5653 (1989). For additional epitopes on Yersinia pastis see caf1 encoding caf1 capsular antigen fraction 1 (F1) accession number X61996 in Galyov et al., FEBS Lett. 277: 230-2 (1990). 12G12 SLP-M common specific epitope of the Brucella S-LPS. S-LPS of B. melitensis Rev1, S-LPS smooth lipopolysaccharide, (Weynants et al., Clin. Diagn. Lab. Immunol. 3:309-14 (1996). Smooth Brucella species – two lipopolysacchiaride (LPSs (A and M) produce monoclonal antibodies to the specific A- or M-LPS epitopes. 12AE6, and Mab 33.1.5, s-LPSA=12AE6, LPSM=33.1.5, BmE10-5, 12G12 O side chain = BRU-38 (Douglas et al., J. Clin. Microbiol. 26: 1353-6 (1988). BmE10-5 Mantigen of Brucella melitensis 16M, B. melitensis, B. abortus, B. suis and B. neotomae (Vizcaino et al., Res. Microbiol. 143: 513-8 (1992). BRU-38, O side chain of B. abortus (Roop et al., J. Clin. Microbiol. 25: 2090-3 (1987). The OMP-specific monoclonal antibodies Bhatti et al., Hybridoma 12: 197-202 (1993). Mabs FT14 and FT2F11, Francisella tularensis (Fulop et al., J. Clin. Microbiol. 29: 1407-12 (1991).

Preferred proteins and multimeric proteins to be made with the present invention include those associated with disease processes and those associated with antigenic epitopes of infectious organisms.

The proteins and multimeric proteins made using the present invention may be labeled using labels and techniques known to one of ordinary skill in the art. Some of these labels are described in the "Handbook of Fluorescent Probes and Research Products", ninth edition, Richard P. Haugland (ed) Molecular Probes, Inc. Eugene, OR), which is incorporated herein in its entirety. Some of these labels may be genetically engineered into the polynucleotide sequence for the expression of the selected multimeric protein. The peptides and proteins may also have label-incorporation "handles" incorporated to allow labeling of an otherwise difficult or impossible to label multimeric protein.

It is to be understood that the various classes of desired peptides and proteins, as well as specific peptides and proteins described in this section may be modified as

described below by inserting selected codons for desired amino acid substitutions into the gene incorporated into the transgenic animal.

The present invention further encompasses the use of inhibitory molecules to inhibit endogenous (i.e., non-vector) protein production. These inhibitory molecules include antisense nucleic acids, siRNA and inhibitory proteins. Preferred inhibitory molecules are those associated with causation or inhibition of disease processes. In a preferred embodiment, the endogenous protein whose expression is inhibited is an egg white protein including, but not limited to ovalbumin, ovotransferrin, and ovomucin ovomucoid, ovoinhibitor, cystatin, ovostatin, lysozyme, ovoglobulin G2, ovoglobulin G3, avidin, and thiamin binding protein. In one embodiment, a polynucleotide cassette containing an ovalbumin DNA sequence, that upon transcription forms a double stranded RNA molecule, is transfected into an animal such as a bird and the bird's production of endogenous ovalbumin protein is reduced by the interference RNA mechanism (RNAi). In other embodiments, a polynucleotide cassette encodes an inhibitory RNA molecule that inhibits the expression of more than one egg white protein. Additionally, inducible knockouts or knockdowns of the endogenous protein may be created to achieve a reduction or inhibition of endogenous protein production. Endogenous egg white production can be inhibited in an avian at any time, but is preferably inhibited preceding, or immediately preceding, the harvest of eggs.

Modified Desired Proteins and Peptides

The present invention may be used for the production of proteins and multimeric proteins. "Proteins", "peptides," "polypeptides" and "oligopeptides" are chains of amino acids (typically L-amino acids) whose alpha carbons are linked through peptide bonds formed by a condensation reaction between the carboxyl group of the alpha carbon of one amino acid and the amino group of the alpha carbon of another amino acid. The terminal amino acid at one end of the chain (i.e., the amino terminal) has a free amino group, while the terminal amino acid at the other end of the chain (i.e., the carboxy terminal) has a free carboxyl group. As such, the term "amino terminus" (abbreviated N-terminus) refers to the free alpha-amino group on the amino acid at the amino terminal of the protein, or to the alpha-amino group (imino group when participating in a peptide bond) of an amino acid at any other location within the protein. Similarly, the term "carboxy terminus" (abbreviated C-terminus) refers to

the free carboxyl group on the amino acid at the carboxy terminus of a protein, or to the carboxyl group of an amino acid at any other location within the protein.

Typically, the amino acids making up a protein are numbered in order, starting at the amino terminal and increasing in the direction toward the carboxy terminal of the protein. Thus, when one amino acid is said to "follow" another, that amino acid is positioned closer to the carboxy terminal of the protein than the preceding amino acid.

The term "residue" is used herein to refer to an amino acid (D or L) or an amino acid mimetic that is incorporated into a protein by an amide bond. As such, the amino acid may be a naturally occurring amino acid or, unless otherwise limited, may encompass known analogs of natural amino acids that function in a manner similar to the naturally occurring amino acids (i.e., amino acid mimetics). Moreover, an amide bond mimetic includes peptide backbone modifications well known to those skilled in the art.

Furthermore, one of skill will recognize that, as mentioned above, individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids (typically less than about 5%, more typically less than about 1%) in an encoded sequence are conservatively modified variations where the alterations result in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. The following six groups each contain amino acids that are conservative substitutions for one another:

1) Alanine (A), Serine (S), Threonine (T);

2) Aspartic acid (D), Glutamic acid (E);

3) Asparagine (N), Glutamine (Q);

4) Arginine (R), Lysine (K);

5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and

6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

A conservative substitution is a substitution in which the substituting amino acid (naturally occurring or modified) is structurally related to the amino acid being substituted, i.e., has about the same size and electronic properties as the amino acid being substituted. Thus, the substituting amino acid would have the same or a similar functional group in the side chain as the original amino acid. A "conservative substitution" also refers to utilizing a substituting amino acid which is identical to the

amino acid being substituted except that a functional group in the side chain is protected with a suitable protecting group.

Suitable protecting groups are described in Green and Wuts, "Protecting Groups in Organic Synthesis", John Wiley and Sons, Chapters 5 and 7, 1991, the teachings of which are incorporated herein by reference. Preferred protecting groups are those which facilitate transport of the peptide through membranes, for example, by reducing the hydrophilicity and increasing the lipophilicity of the peptide, and which can be cleaved, either by hydrolysis or enzymatically (Ditter et al., 1968. J. Pharm. Sci. 57:783; Ditter et al., 1968. J. Pharm. Sci. 57:828; Ditter et al., 1969. J. Pharm. Sci. 58:557; King et al., 1987. Biochemistry 26:2294; Lindberg et al., 1989. Drug Metabolism and Disposition 17:311; Tunek et al., 1988. Biochem. Pharm. 37:3867; Anderson et al., 1985 Arch. Biochem. Biophys. 239:538; and Singhal et al., 1987. FASEB J. 1:220). Suitable hydroxyl protecting groups include ester, carbonate and carbamate protecting groups. Suitable amine protecting groups include acyl groups and alkoxy or aryloxy carbonyl groups, as described above for N-terminal protecting groups. Suitable carboxylic acid protecting groups include aliphatic, benzyl and aryl esters, as described below for C-terminal protecting groups. In one embodiment, the carboxylic acid group in the side chain of one or more glutamic acid or aspartic acid residues in a peptide of the present invention is protected, preferably as a methyl, ethyl, benzyl or substituted benzyl ester, more preferably as a benzyl ester.

Provided below are groups of naturally occurring and modified amino acids in which each amino acid in a group has similar electronic and steric properties. Thus, a conservative substitution can be made by substituting an amino acid with another amino acid from the same group. It is to be understood that these groups are non-limiting, i.e. that there are additional modified amino acids which could be included in each group.

Group I includes leucine, isoleucine, valine, methionine and modified amino acids having the following side chains: ethyl, n-propyl n-butyl. Preferably, Group I includes leucine, isoleucine, valine and methionine.

Group II includes glycine, alanine, valine and a modified amino acid having an ethyl side chain. Preferably, Group II includes glycine and alanine.

Group III includes phenylalanine, phenylglycine, tyrosine, tryptophan, cyclohexylmethyl glycine, and modified amino residues having substituted

benzyl or phenyl side chains. Preferred substituents include one or more of the following: halogen, methyl, ethyl, nitro, —NH_2, methoxy, ethoxy and —CN. Preferably, Group III includes phenylalanine, tyrosine and tryptophan.

Group IV includes glutamic acid, aspartic acid, a substituted or unsubstituted aliphatic, aromatic or benzylic ester of glutamic or aspartic acid (e.g., methyl, ethyl, n-propyl iso-propyl, cyclohexyl, benzyl or substituted benzyl), glutamine, asparagine, —CO—NH— alkylated glutamine or asparagines (e.g., methyl, ethyl, n-propyl and iso-propyl) and modified amino acids having the side chain —$(CH_2)_3$—COOH, an ester thereof (substituted or unsubstituted aliphatic, aromatic or benzylic ester), an amide thereof and a substituted or unsubstituted N-alkylated amide thereof. Preferably, Group IV includes glutamic acid, aspartic acid, methyl aspartate, ethyl aspartate, benzyl aspartate and methyl glutamate, ethyl glutamate and benzyl glutamate, glutamine and asparagine.

Group V includes histidine, lysine, ornithine, arginine, N-nitroarginine, ß-cycloarginine, γ-hydroxyarginine, N-amidinocitruline and 2-amino-4-guanidinobutanoic acid, homologs of lysine, homologs of arginine and homologs of ornithine. Preferably, Group V includes histidine, lysine, arginine and ornithine. A homolog of an amino acid includes from 1 to about 3 additional or subtracted methylene units in the side chain.

Group VI includes serine, threonine, cysteine and modified amino acids having C1-C5 straight or branched alkyl side chains substituted with —OH or —SH, for example, —CH_2CH_2OH, —$CH_2CH_2CH_2OH$ or -$CH_2CH_2OHCH_3$. Preferably, Group VI includes serine, cysteine or threonine.

In another aspect, suitable substitutions for amino acid residues include "severe" substitutions. A "severe substitution" is a substitution in which the substituting amino acid (naturally occurring or modified) has significantly different size and/or electronic properties compared with the amino acid being substituted. Thus, the side chain of the substituting amino acid can be significantly larger (or smaller) than the side chain of the amino acid being substituted and/or can have functional groups with significantly different electronic properties than the amino acid being substituted. Examples of severe substitutions of this type include the substitution of phenylalanine or cyclohexylmethyl glycine for alanine, isoleucine for

glycine, a D amino acid for the corresponding L amino acid, or —NH—CH[(—$CH_2)_5$—COOH]—CO— for aspartic acid. Alternatively, a functional group may be added to the side chain, deleted from the side chain or exchanged with another functional group. Examples of severe substitutions of this type include adding of valine, leucine or isoleucine, exchanging the carboxylic acid in the side chain of aspartic acid or glutamic acid with an amine, or deleting the amine group in the side chain of lysine or ornithine. In yet another alternative, the side chain of the substituting amino acid can have significantly different steric and electronic properties that the functional group of the amino acid being substituted. Examples of such modifications include tryptophan for glycine, lysine for aspartic acid and —$(CH_2)_4$COOH for the side chain of serine. These examples are not meant to be limiting.

In another embodiment, for example in the synthesis of a peptide 26 amino acids in length, the individual amino acids may be substituted according in the following manner:

AA_1 is serine, glycine, alanine, cysteine or threonine;

AA_2 is alanine, threonine, glycine, cysteine or serine;

AA_3 is valine, arginine, leucine, isoleucine, methionine, ornithine, lysine, N-nitroarginine, ß-cycloarginine, γ-hydroxyarginine, N-amidinocitruline or 2-amino-4-guanidinobutanoic acid;

AA_4 is proline, leucine, valine, isoleucine or methionine;

AA_5 is tryptophan, alanine, phenylalanine, tyrosine or glycine;

AA_6 is serine, glycine, alanine, cysteine or threonine;

AA_7 is proline, leucine, valine, isoleucine or methionine;

AA_8 is alanine, threonine, glycine, cysteine or serine;

AA_9 is alanine, threonine, glycine, cysteine or serine;

AA_{10} is leucine, isoleucine, methionine or valine;

AA_{11} is serine, glycine, alanine, cysteine or threonine;

AA_{12} is leucine, isoleucine, methionine or valine;

AA_{13} is leucine, isoleucine, methionine or valine;

AA_{14} is glutamine, glutamic acid, aspartic acid, asparagine, or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;

AA_{15} is arginine, N-nitroarginine, ß-cycloarginine, γ-hydroxy-arginine, N-amidinocitruline or 2-amino-4-guanidino-butanoic acid

AA_{16} is proline, leucine, valine, isoleucine or methionine;

AA_{17} is serine, glycine, alanine, cysteine or threonine;

AA_{18} is glutamic acid, aspartic acid, asparagine, glutamine or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;

AA_{19} is aspartic acid, asparagine, glutamic acid, glutamine, leucine, valine, isoleucine, methionine or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;

AA_{20} is valine, arginine, leucine, isoleucine, methionine, ornithine, lysine, N-nitroarginine, ß-cycloarginine, γ-hydroxyarginine, N-amidinocitruline or 2-amino-4-guanidinobutanoic acid;

AA_{21} is alanine, threonine, glycine, cysteine or serine;

AA_{22} is alanine, threonine, glycine, cysteine or serine;

AA_{23} is histidine, serine, threonine, cysteine, lysine or ornithine;

AA_{24} is threonine, aspartic acid, serine, glutamic acid or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;

AA_{25} is asparagine, aspartic acid,, glutamic acid, glutamine, leucine, valine, isoleucine, methionine or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid; and

AA_{26} is cysteine, histidine, serine, threonine, lysine or ornithine.

It is to be understood that these amino acid substitutions may be made for longer or shorter peptides than the 26 mer in the preceding example above, and for proteins.

In one embodiment of the present invention, codons for the first several N-terminal amino acids of the transposase are modified such that the third base of each codon is changed to an A or a T without changing the corresponding amino acid. It is preferable that between approximately 1 and 20, more preferably 3 and 15, and most preferably between 4 and 12 of the first N-terminal codons of the gene of interest are modified such that the third base of each codon is changed to an A or a T without changing the corresponding amino acid. In one embodiment, the first ten N-terminal codons of the gene of interest are modified in this manner.

When several desired proteins, protein fragments or peptides are encoded in the gene of interest to be incorporated into the genome, as with the multivalent multimeric proteins, one of skill in the art will appreciate that the proteins, protein fragments or peptides may be separated by a spacer molecule such as, for example, a peptide, consisting of one or more amino acids. Generally, the spacer will have no specific biological activity other than to join the desired proteins, protein fragments or peptides together, or to preserve some minimum distance or other spatial relationship between them. However, the constituent amino acids of the spacer may be selected to influence some property of the molecule such as the folding, net charge, or hydrophobicity. The spacer may also be contained within a nucleotide sequence with a purification handle or be flanked by proteolytic cleavage sites.

Such polypeptide spacers may have from about 1 to about 100 amino acids, preferably 3 to 20 amino acids, and more preferably 4-15 amino acids. The spacers in a polypeptide are independently chosen, but are preferably all the same. The spacers should allow for flexibility of movement in space and are therefore typically rich in small amino acids, for example, glycine, serine, proline or alanine. Preferably, peptide spacers contain at least 60%, more preferably at least 80% glycine or alanine. In addition, peptide spacers generally have little or no biological and antigenic activity. Preferred spacers are $(Gly\text{-}Pro\text{-}Gly\text{-}Gly)_x$ (SEQ ID NO:81) and $(Gly_4\text{-}Ser)_y$, wherein x is an integer from about 3 to about 9 and y is an integer from about 1 to about 8. Specific examples of suitable spacers include

$(Gly\text{-}Pro\text{-}Gly\text{-}Gly)_3$

SEQ ID NO:82 Gly Pro Gly Gly Gly Pro Gly Gly Gly Pro Gly Gly

$(Gly_4\text{-}Ser)_3$

SEQ ID NO:83 Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser

or $(Gly_4\text{-}Ser)_4$

SEQ ID NO:84 Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser.

One example of a multivalent multimeric protein containin a spacer is leutinizing hormone (LH), normally made as separate alpha and beta chains, made as a single polypeptide as described in Galet et. al., Mol. Cell Endocrinology, 2001, 174 (1-2):31-40. Production of a multimeric protein may thus be simplified using a spacer sequence that may or may not contain cleavage sites. In the case of an

immunoglobulin, for example, a heavy and light chain may be synthesized as a single polypeptide using a spacer sequence with protease sites native to the transgenic animal so as to make, upon processing, a heavy and light chain combination in close association, facilitating the addition of a similar heavy and light chain to produce the native immunoglobulin. In this model, the removal of the spacer sequence may or may not be required. Other multimeric proteins may be made in bioengineered organisms in a similar fashion.

Nucleotide sequences encoding for the production of residues which may be useful in purification of the expressed recombinant protein may also be built into the vector. Such sequences are known in the art and include the glutathione binding domain from glutathione S-transferase, polylysine, hexa-histidine or other cationic amino acids, thioredoxin, hemagglutinin antigen and maltose binding protein.

Additionally, nucleotide sequences may be inserted into the gene of interest to be incorporated so that the protein or peptide can also include from one to about six amino acids that create signals for proteolytic cleavage. In this manner, if a gene is designed to make one or more peptides or proteins of interest in the transgenic animal, specific nucleotide sequences encoding for amino acids recognized by enzymes may be incorporated into the gene to facilitate cleavage of the large protein or peptide sequence into desired peptides or proteins or both. For example, nucleotides encoding a proteolytic cleavage site can be introduced into the gene of interest so that a signal sequence can be cleaved from a protein or peptide encoded by the gene of interest. Nucleotide sequences encoding other amino acid sequences which display pH sensitivity, chemical sensitivity or photolability may also be added to the vector to facilitate separation of the signal sequence from the peptide or protein of interest.

Proteolytic cleavage sites include cleavage sites recognized by exopeptidases such as carboxypeptidase A, carboxypeptidase B, aminopeptidase I, and dipeptidylaminopeptidase; endopeptidases such as trypsin, V8-protease, enterokinase, factor Xa, collagenase, endoproteinase, subtilisin, and thombin; and proteases such as Protease 3C IgA protease (Igase) Rhinovirus 3C(preScission)protease. Chemical cleavage sites are also included in the defintion of cleavage site as used herein. Chemical cleavage sites include, but are not limited to, site cleaved by cyanogen bromide, hydroxylamine, formic acid, and acetic acid. Self-splicing cleavage sites such as inteins are also included in the present invention.

In some embodiments, one or more cleavage sites are incorporated into a polynucleotide cassette containing multiple genes of interest. Figure 4 depicts one example of a polynucleotide cassette containing two genes of interest containing a cleavage site between them. The genes of interest may encode different proteins or peptides, the same protein or peptide, or modified versions of the same protein or peptide. While Figure 4 shows a polynucleotide cassette containing two genes of interest, the present invention encompasses a polynucleotide cassette containing any number of genes of interest. The cleavage site located between the genes of interest can encode any amino acid sequence that is cleaved by any means. As mentioned above, the cleavage site can encode an amino acid sequence cleaved by a protease, a chemical reaction, can be a photolabile site, or can be a pro polynucleotide.

The present invention includes a polynucleotide cassette that encodes a repetitive polypeptide chain in which two or more peptides, polypeptides or proteins, designated as P in the structural formulae presented below, are each separated by a peptide spacer or cleavage site designated as B. A polypeptide multivalent ligand, also called a multivalent protein, is a form of a multimeric protein encoded by the polynucleotide cassettes of the present invention, and is represented by structural formulae (I, II and III). Each peptide or protein is connected to another peptide or protein through a peptide bond, to a linker group, to a spacer, or to a cleavage site. Each peptide, polypeptide or protein may be the same or different and each linker, spacer, cleavage site or covalent bond is independently chosen.

A "polypeptide multivalent protein" is a multiple repeat polypeptide chain in which two or more peptides P are each separated by a peptide linker group, a spacer or a cleavage site. A polypeptide multivalent ligand is represented by structural formulae II and III.

I B—(L—P)n

wherein B is a peptide spacer or cleavage site, n is an integer from 2 to about 20, each L is a covalent bond, a linking group or cleavage site which may be present or absent, and each P is a peptide having from about 4 to about 200 amino acid residues.

II P—(B—P)m—B—P

wherein m is an integer from 0 to about 20.

III $Pa-(B)n-Pa$

wherein n is an integer from 1 to 20, preferably 2 to 10, more preferably 3 to 7, further wherein a is 1.

Other examples of multivalent proteins include the following:

IV $P_y——L_x——B_n——L_x——P_y$

V $P_y——B——P_y$

In the preceding structural formulae IV and V of polypeptide multivalent ligands encoded by a polynucleotide cassette of the present invention, each P is a peptide having from about 4 to about 200 amino acid residues, y is 1, x is an integer from 1 to 3, and n is an integer from 1 to 20, preferably 2 to 10, more preferably 3 to 7. Each B is a peptide spacer or cleavage site comprised of at least 2 amino acids or a cleavage site. Each peptide P and each B are independently chosen and may be the same or different.

Suitable linkers (L) are groups that can connect peptides and proteins to each other. In one example, the linker is an oligopeptide of from about 1 to about 10 amino acids consisting of amino acids with inert side chains. Suitable oligopeptides include polyglycine, polyserine, polyproline, polyalanine and oligopeptides consisting of alanyl and/or serinyl and/or prolinyl and/or glycyl amino acid residues. m in structural formula II is an integer from 0 to about 20.

The peptides, polypeptides and proteins in a multivalent protein can be connected to each other by covalent bonds, linker groups, spacers, cleavage groups or a combination thereof. The linking groups can be the same or different.

A polypeptide spacer shown in structural formula (II) is a peptide having from about 5 to about 40 amino acid residues. The spacers in a polypeptide multivalent ligand are independently chosen, and may be the same or different. The spacers should allow for flexibility of movement in space for the flanking peptides. polypeptides and proteins P, and are therefore typically rich in small amino acids, for

example, glycine, serine, proline or alanine. Preferably, peptide spacers contain at least 60%, more preferably at least 80% glycine or alanine. In addition, peptide spacers generally have little or no biological and antigenic activity. Preferred spacers are $(Gly\text{-}Pro\text{-}Gly\text{-}Gly)_x$ (SEQ ID NO:81) and $(Gly_4\text{-}Ser)_y$, wherein x is an integer from about 3 to about 9 and y is an integer from about 1 to about 8. Specific examples of suitable spacers include $(Gly_4\text{-}Ser)_3$ (SEQ ID NO:82). Spacers can also include from one to about four amino acids that create a signal for proteolytic cleavage.

In another embodiment of the present invention, a TAG sequence is linked to a gene of interest. The TAG sequence serves three purposes: 1) it allows free rotation of the peptide or protein to be isolated so there is no interference from the native protein or signal sequence, i.e. vitellogenin, 2) it provides a "purification handle" to isolate the protein using affinity purification, and 3) it includes a cleavage site to remove the desired protein from the signal and purification sequences. Accordingly, as used herein, a TAG sequence includes a spacer sequence, a purification handle and a cleavage site. The spacer sequences in the TAG proteins contain one or more repeats shown in SEQ ID NO:85. A preferred spacer sequence comprises the sequence provided in SEQ ID NO:86. One example of a purification handle is the gp41 hairpin loop from HIV I. Exemplary gp41 polynucleotide and polypeptide sequences are provided in SEQ ID NO:87 and SEQ ID NO:88, respectively. However, it should be understood that any antigenic region, or otherwise associative regions such as avidin/biotin, may be used as a purification handle, including any antigenic region of gp41. Preferred purification handles are those that elicit highly specific antibodies. Additionally, the cleavage site can be any protein cleavage site known to one of ordinary skill in the art and includes an enterokinase cleavage site comprising the Asp Asp Asp Asp Lys sequence (SEQ ID NO:89) and a furin cleavage site. In one embodiment of the present invention, the TAG sequence comprises a polynucleotide sequence of SEQ ID NO:90.

Methods of Administering Polynucleotide Cassettes

In addition to the polynucleotide cassettes described above, the present invention also includes methods of administering the polynucleotide cassettes to an animal, methods of producing a transgenic animal wherein a gene of interest is incorporated into the germline of the animal and methods of producing a transgenic

animal wherein a gene of interest is incorporated into cells other than the germline cells of the animal. The polynucleotide cassettes may reside in any vector or delivery solution when administered or may be naked DNA. In one embodiment, a transposon-based vector containing the polynucleotide cassette between two insertion sequences recognized by a transposase is administered to an animal. The polynucleotide cassettes of the present invention may be administered to an animal via any method known to those of skill in the art, including, but not limited to, intraembryonic, intratesticular, intraoviduct, intraovarian, into the duct system of the mammary gland, intraperitoneal, intraarterial, intravenous, topical, oral, nasal, and pronuclear injection methods of administration, or any combination thereof. The polynucleotide cassettes may also be administered within the lumen of an organ, into an organ, into a body cavity, into the cerebrospinal fluid, through the urinary system, through the genitourinary system, through the reproductive system, or through any route to reach the desired cells.

The polynucleotide cassettes may be delivered through the vascular system to be distributed to the cells supplied by that vessel. For example, the compositions may be placed in the artery supplying the ovary or supplying the fallopian tube to transfect cells in those tissues. In this manner, follicles could be transfected to create a germline transgenic animal. Alternatively, supplying the compositions through the artery leading to the oviduct would preferably transfect the tubular gland and epithelial cells. Such transfected cells could manufacture a desired protein or peptide for deposition in the egg white. Administration of the compositions through the portal vein would target uptake and transformation of hepatic cells. Administration through the urethra and into the bladder would target the transitional epithelium of the bladder. Administration through the vagina and cervix would target the lining of the uterus and the epithelial cells of the fallopian tube. Administration through the internal mammary artery or through the duct system of the mammary gland would transfect secretory cells of the lactating mammary gland to perform a desired function, such as to synthesize and secrete a desired protein or peptide into the milk.

The polynucleotide cassettes may be administered in a single administration, multiple administrations, continuously, or intermittently. The polynucleotide cassettes may be administered by injection, via a catheter, an osmotic mini-pump or any other method. In some embodiments, a polynucleotide cassette is administered to

an animal in multiple administrations, each administration containing the polynucleotide cassette and a different transfecting reagent.

In a preferred embodiment, the animal is an egg-laying animal, and more preferably, an avian. In one embodiment, between approximately 1 and 150 μg, 1 and 100 μg, 1 and 50 μg, preferably between 1 and 20 μg, and more preferably between 5 and 10 μgof a transposon-based vector containing the polynucleotide cassette is administered to the oviduct of a bird. In a chicken, it is preferred that between approximately 1 and 100 μg, or 5 and 50 μg are administered. In a quail, it is preferred that between approximately 5 and 10 μg are administered. Optimal ranges depending upon the type of bird and the bird's stage of sexual maturity. Intraoviduct administration of the transposon-based vectors of the present invention result in a PCR positive signal in the oviduct tissue, whereas intravascular administration results in a PCR positive signal in the liver. In other embodiments, the polynucleotide cassettes is administered to an artery that supplies the oviduct or the liver. These methods of administration may also be combined with any methods for facilitating transfection, including without limitation, electroporation, gene guns, injection of naked DNA, and use of dimethyl sulfoxide (DMSO).

The transposon-based vectors may be administered to the animal at any point during the lifetime of the animal, however, it is preferable that the vectors are administered prior to the animal reaching sexual maturity. The transposon-based vectors are preferably administered to a chicken oviduct between approximately 14 and 16 weeks of age and to a quail oviduct between approximately 5 and 10 weeks of age, more preferably 5 and 8 weeks of age, and most preferably between 5 and 6 weeks of age, when standard poultry rearing practices are used. The vectors may be administered at earlier ages when exogenous hormones are used to induce early sexual maturation in the bird. In some embodiments, the transposon-based vector is administered to an animal's oviduct following an increase in proliferation of the oviduct epithelial cells and/or the tubular gland cells. Such an increase in proliferation normally follows an influx of reproductive hormones in the area of the oviduct. When the animal is an avian, the transposon-based vector is administered to the avian's oviduct following an increase in proliferation of the oviduct epithelial cells and before the avian begins to produce egg white constituents.

The present invention also includes a method of intraembryonic administration of a transposon-based vector containing a polynucleotide cassette to an avian embryo comprising the following steps: 1) incubating an egg on its side at room temperature for two hours to allow the embryo contained therein to move to top dead center (TDC); 2) drilling a hole through the shell without penetrating the underlying shell membrane; 3) injecting the embryo with the transposon-based vector in solution; 4) sealing the hole in the egg; and 5) placing the egg in an incubator for hatching. Administration of the transposon-based vector can occur anytime between immediately after egg lay (when the embryo is at Stage X) and hatching. Preferably, the transposon-based vector is administered between 1 and 7 days after egg lay, more preferably between 1 and 2 days after egg lay. The transposon-based vectors may be introduced into the embryo in amounts ranging from about 5.0 μg to 10 pg, preferably 1.0 μg to 100 pg. Additionally, the transposon-based vector solution volume may be between approximately 1 μl to 75 μl in quail and between approximately 1 μl to 500 μl in chicken.

The present invention also includes a method of intratesticular administration of a transposon-based vector containing a polynucleotide cassette including injecting a bird with a composition comprising the transposon-based vector, an appropriate carrier and an appropriate transfection reagent. In one embodiment, the bird is injected before sexual maturity, preferably between approximately 4-14 weeks, more preferably between approximately 6-14 weeks and most preferably between 8-12 weeks old. In another embodiment, a mature bird is injected with a transposon-based vector an appropriate carrier and an appropriate transfection reagent. The mature bird may be any type of bird, but in one example the mature bird is a quail.

A bird is preferably injected prior to the development of the blood-testis barrier, which thereby facilitates entry of the transposon-based vector into the seminiferous tubules and transfection of the spermatogonia or other germline cells. At and between the ages of 4, 6, 8, 10, 12, and 14 weeks, it is believed that the testes of chickens are likely to be most receptive to transfection. In this age range, the blood/testis barrier has not yet formed, and there is a relatively high number of spermatogonia relative to the numbers of other cell types, e.g., spermatids, etc. See J. Kumaran et al., 1949. Poultry Sci., 29:511-520. See also E. Oakberg, 1956. Am. J. Anatomy, 99:507-515; and P. Kluin et al., 1984. Anat. Embryol., 169:73-78.

The transposon-based vectors may be introduced into a testis in an amount ranging from about 0.1 μg to 10 μg, preferably 1 μg to 10 μg, more preferably 3 μg to 10 μg. In a quail, about 5 μg is a preferred amount. In a chicken, about 5 μg to 10 μg per testis is preferred. These amounts of vector DNA may be injected in one dose or multiple doses and at one site or multiple sites in the testis. In a preferred embodiment, the vector DNA is administered at multiple sites in a single testis, both testes being injected in this manner. In one embodiment, injection is spread over three injection sites: one at each end of the testis, and one in the middle. Additionally, the transposon-based vector solution volume may be between approximately 1 μl to 75 μl in quail and between approximately 1 μl to 500 μl in chicken. In a preferred embodiment, the transposon-based vector solution volume may be between approximately 20 μl to 60 μl in quail and between approximately 50 μl to 250 μl in chicken. Both the amount of vector DNA and the total volume injected into each testis may be determined based upon the age and size of the bird.

According to the present invention, the polynucleotide cassette is administered in conjunction with an acceptable carrier and/or transfection reagent. Acceptable carriers include, but are not limited to, water, saline, Hanks Balanced Salt Solution (HBSS), Tris-EDTA (TE) and lyotropic liquid crystals. Transfection reagents commonly known to one of ordinary skill in the art that may be employed include, but are not limited to, the following: cationic lipid transfection reagents, cationic lipid mixtures, polyamine reagents, liposomes and combinations thereof; SUPERFECT®, Cytofectene, BioPORTER®, GenePORTER®, NeuroPORTER®, and perfectin from Gene Therapy Systems; lipofectamine, cellfectin, DMRIE-C oligofectamine, and PLUS reagent from InVitrogen; Xtreme gene, fugene, DOSPER and DOTAP from Roche; Lipotaxi and Genejammer from Strategene; and Escort from SIGMA. In one embodiment, the transfection reagent is SUPERFECT®. The ratio of DNA to transfection reagent may vary based upon the method of administration. In one embodiment, a transposon-based vector containing a polynucleotide cassette is administered intratesticularly and the ratio of DNA to transfection reagent can be from 1:1.5 to 1:15, preferably 1:2 to 1:10, all expressed as wt/vol. Transfection may also be accomplished using other means known to one of ordinary skill in the art, including without limitation electroporation, gene guns, injection of naked DNA, and use of dimethyl sulfoxide (DMSO).

Depending upon the cell or tissue type targeted for transfection, the form of the transposon-based vector may be important. Plasmids harvested from bacteria are generally closed circular supercoiled molecules, and this is the preferred state of a vector for gene delivery because of the ease of preparation. In some instances, transposase expression and insertion may be more efficient in a relaxed, closed circular configuration or in a linear configuration. In still other instances, a purified transposase protein may be co-injected with a transposon-based vector containing the gene of interest for more immediate insertion. This could be accomplished by using a transfection reagent complexed with both the purified transposase protein and the transposon-based vector.

Testing for and Breeding Animals Carrying the Transgene

Following administration of a polynucleotide cassette to an animal, DNA is extracted from the animal to confirm integration of the genes of interest. Advantages provided by the present invention include the high rates of integration, or incorporation, and transcription of the gene of interest when administered to a bird via an intraoviduct or intraovary route (including intraarterial administrations to arteries leading to the oviduct or ovary) and contained within a transposon-based vector.

Actual frequencies of integration can be estimated both by comparative strength of the PCR signal, and by histological evaluation of the tissues by quantitative PCR. Another method for estimating the rate of transgene insertion is the so-called primed in situ hybridization technique (PRINS). This method determines not only which cells carry a transgene of interest, but also into which chromosome the gene has inserted, and even what portion of the chromosome. Briefly, labeled primers are annealed to chromosome spreads (affixed to glass slides) through one round of PCR, and the slides are then developed through normal in situ hybridization procedures. This technique combines the best features of in situ PCR and fluorescence in situ hybridization (FISH) to provide distinct chromosome location and copy number of the gene in question. The 28s rRNA gene will be used as a positive control for spermatogonia to confirm that the technique is functioning properly. Using different fluorescent labels for the transgene and the 28s gene causes cells containing a transgene to fluoresce with two different colored tags.

Breeding experiments may also be conducted to determine if germline transmission of the transgene has occurred. In a general bird breeding experiment

performed according to the present invention, each male bird is exposed to 2-3 different adult female birds for 3-4 days each. This procedure is continued with different females for a total period of 6-12 weeks. Eggs are collected daily for up to 14 days after the last exposure to the transgenic male, and each egg is incubated in a standard incubator. The resulting embryos are examined for transgene presence at day 3 or 4 using PCR.

Any male producing a transgenic embryo is bred to additional females. Eggs from these females are incubated, hatched, and the chicks tested for the exogenous DNA. Any embryos that die are necropsied and examined directly for the transgene or protein encoded by the transgene, either by fluorescence or PCR. The offspring that hatch and are found to be positive for the exogenous DNA are raised to maturity. These birds are bred to produce further generations of transgenic birds, to verify efficiency of the transgenic procedure and the stable incorporation of the transgene into the germ line. The resulting embryos are examined for transgene presence at day 3 or 4 using PCR.

It is to be understood that the above procedure can be modified to suit animals other than birds and that selective breeding techniques may be performed to amplify gene copy numbers and protein output.

Production of Desired Proteins and Multimeric Proteins in Egg White

In one embodiment, a transposon-based vector containing a polynucleotide cassette of the present invention may be administered to a bird for production of desired proteins or peptides in the egg white. These trasnposon-based vectors preferably contain one or more of an ovalbumin promoter, an ovomucoid promoter, an ovalbumin signal sequence and an ovomucoid signal sequence. Oviduct-specific ovalbumin promoters are described in B. O'Malley et al., 1987. EMBO J., vol. 6, pp. 2305-12; A. Qiu et al., 1994. Proc. Nat. Acad. Sci. (USA), vol. 91, pp. 4451-4455; D. Monroe et al., 2000. Biochim. Biophys. Acta, 1517 (1):27-32; H. Park et al., 2000. Biochem., 39:8537-8545; and T. Muramatsu et al., 1996. Poult. Avian Biol. Rev., 6:107-123.

Production of Desired Proteins and Multimeric Proteins in Egg Yolk

The present invention is particularly advantageous for production of recombinant peptides and proteins of low solubility in the egg yolk. Such proteins include, but are not limited to, membrane-associated or membrane-bound proteins,

lipophilic compounds; attachment factors, receptors, and components of second messenger transduction machinery. Low solubility peptides and proteins are particularly challenging to produce using conventional recombinant protein production techniques (cell and tissue cultures) because they aggregate in water-based, hydrophilic environments. Such aggregation necessitates denaturation and re-folding of the recombinantly-produced proteins, which may deleteriously affect their structure and function. Moreover, even highly soluble recombinant peptides and proteins may precipitate and require denaturation and renaturation when produced in sufficiently high amounts in recombinant protein production systems. The present invention provides an advantageous resolution of the problem of protein and peptide solubility during production of large amounts of recombinant proteins.

In one embodiment of the present invention, deposition of a desired protein into the egg yolk is accomplished by attaching a sequence encoding a protein capable of binding to the yolk vitellogenin receptor to a gene of interest that encodes a desired protein. This polynucleotide cassette can be used for the receptor-mediated uptake of the desired protein by the oocytes. In a preferred embodiment, the sequence ensuring the binding to the vitellogenin receptor is a targeting sequence of a vitellogenin protein. The invention encompasses various vitellogenin proteins and their targeting sequences. In a preferred embodiment, a chicken vitellogenin protein targeting sequence is used, however, due to the high degree of conservation among vitellogenin protein sequences and known cross-species reactivity of vitellogenin targeting sequences with their egg-yolk receptors, other vitellogenin targeting sequences can be substituted. One example of a construct for use in the transposon-based vectors of the present invention and for deposition of an insulin protein in an egg yolk is a transposon-based vector containing a vitellogenin promoter, a vitellogenin targeting sequence, a TAG sequence, a pro-insulin sequence and a synthetic polyA sequence. The present invention includes, but is not limited to, vitellogenin targeting sequences residing in the N-terminal domain of vitellogenin, particularly in lipovitellin I. In one embodiment, the vitellogenin targeting sequence contains the polynucleotide sequence of SEQ ID NO:77.

In a preferred embodiment, the transposon-based vector contains a transposase gene operably-linked to a constitutive promoter and a gene of interest operably-linked to a liver-specific promoter and a vitellogenin targeting sequence.

Isolation and Purification of Desired Proteins and Multimeric Proteins

For large-scale production of protein, an animal breeding stock that is homozygous for the transgene is preferred. Such homozygous individuals are obtained and identified through, for example, standard animal breeding procedures or PCR protocols.

Once expressed, peptides, polypeptides and proteins can be purified according to standard procedures known to one of ordinary skill in the art, including ammonium sulfate precipitation, affinity columns, column chromatography, gel electrophoresis, high performance liquid chromatography, immunoprecipitation and the like. Substantially pure compositions of about 50 to 99% homogeneity are preferred, and 80 to 95% or greater homogeneity are most preferred for use as therapeutic agents.

In one embodiment of the present invention, the animal in which the desired protein is produced is an egg-laying animal. In a preferred embodiment of the present invention, the animal is an avian and a desired peptide, polypeptide or protein is isolated from an egg white. Egg white containing the exogenous protein or peptide is separated from the yolk and other egg constituents on an industrial scale by any of a variety of methods known in the egg industry. See, e.g., W. Stadelman et al. (Eds.), Egg Science & Technology, Haworth Press, Binghamton, NY (1995). Isolation of the exogenous peptide or protein from the other egg white constituents is accomplished by any of a number of polypeptide isolation and purification methods well known to one of ordinary skill in the art. These techniques include, for example, chromatographic methods such as gel permeation, ion exchange, affinity separation, metal chelation, HPLC, hydrophobic interaction chromatography, and the like, either alone or in combination. Such techniques are known to one of ordinary skill in the art of protein purification. Another means that may be used for isolation or purification, either in lieu of or in addition to chromatographic separation methods, includes electrophoresis. Successful isolation and purification is confirmed by standard analytic techniques, including HPLC, mass spectroscopy, and spectrophotometry. These separation methods are often facilitated if the first step in the separation is the removal of the endogenous ovalbumin fraction of egg white, as doing so will reduce the total protein content to be further purified by about 50%.

To facilitate or enable purification of a desired protein or peptide, the polynucleotide cassettes may include one or more additional epitopes or domains.

Such epitopes or domains include DNA sequences encoding enzymatic, chemical or photolabile cleavage sites including, but not limited to, an enterokinase cleavage site; the glutathione binding domain from glutathione S-transferase; polylysine; hexa-histidine or other cationic amino acids, and sites cleaved by cyanogen bromide, hydroxylamine, formic acid, and acetic acid; thioredoxin; hemagglutinin antigen; maltose binding protein; a fragment of gp41 from HIV; and other purification epitopes or domains commonly known to one of skill in the art. Other proteolytic cleavage sites that may be included in the polynucleotide cassettes are cleavage sites recognized by exopeptidases such as carboxypeptidase A, carboxypeptidase B, aminopeptidase I, and dipeptidylaminopeptidase; endopeptidases such as trypsin, V8-protease, enterokinase, factor Xa, collagenase, endoproteinase, subtilisin, and thombin; and proteases such as Protease 3C IgA protease (Igase) Rhinovirus 3C(preScission)protease. Self-splicing cleavage sites such as inteins may also be included in the polynucleotide cassettes of the present invention.

In one representative embodiment, purification of desired proteins from egg white utilizes the antigenicity of the ovalbumin carrier protein and particular attributes of a TAG linker sequence that spans ovalbumin and the desired protein. The TAG sequence is particularly useful in this process because it contains 1) a highly antigenic epitope, a fragment of gp41 from HIV, allowing for stringent affinity purification, and, 2) a recognition site for the protease enterokinase immediately juxtaposed to the desired protein. In a preferred embodiment, the TAG sequence comprises approximately 50 amino acids. A representative TAG sequence is provided below.

Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp <u>Ala Thr Thr Cys Ile Leu Lys Gly Ser Cys Gly Trp Ile Gly Leu Leu</u> *Asp Asp Asp Asp Lys* (SEQ ID NO:90)

The underlined sequences were taken from the hairpin loop domain of HIV gp-41 (SEQ ID NO:87). Sequences in italics represent the cleavage site for enterokinase (SEQ ID NO:89). The spacer sequence upstream of the loop domain was made from repeats of (Pro Ala Asp Asp Ala) (SEQ ID NO:85) to provide free rotation and promote surface availability of the hairpin loop from the ovalbumin carrier protein.

Isolation and purification of a desired protein is performed as follows:

1. Enrichment of the egg white protein fraction containing ovalbumin and the transgenic ovalbumin-TAG-desired protein.
2. Size exclusion chromatography to isolate only those proteins within a narrow range of molecular weights (a further enrichment of step 1).
3. Ovalbumin affinity chromatography. Highly specific antibodies to ovalbumin will eliminate virtually all extraneous egg white proteins except ovalbumin and the transgenic ovalbumin-TAG-desired protein.
4. gp41 affinity chromatography using anti-gp41 antibodies. Stringent application of this step will result in virtually pure transgenic ovalbumin-TAG-desired protein.
5. Cleavage of the transgene product can be accomplished in at least one of two ways:
    a. The transgenic ovalbumin-TAG-desired protein is left attached to the gp41 affinity resin (beads) from step 4 and the protease enterokinase is added. This liberates the transgene target protein from the gp41 affinity resin while the ovalbumin-TAG sequence is retained. Separation by centrifugation (in a batch process) or flow through (in a column purification), leaves the desired protein together with enterokinase in solution. Enterokinase is recovered and reused.
    b. Alternatively, enterokinase is immobilized on resin (beads) by the addition of poly-lysine moieties to a non-catalytic area of the protease. The transgenic ovalbumin-TAG-desired protein eluted from the affinity column of step 4 is then applied to the protease resin. Protease action cleaves the ovalbumin-TAG sequence from the desired protein and leaves both entities in solution. The immobilized enterokinase resin is recharged and reused.
    c. The choice of these alternatives is made depending upon the size and chemical composition of the transgene target protein.
6. A final separation of either of these two (5a or 5b) protein mixtures is made using size exclusion, or enterokinase affinity chromatography. This step allows for desalting, buffer exchange and/or polishing, as needed.

Cleavage of the transgene product (ovalbumin-TAG-desired protein) by enterokinase, then, results in two products: ovalbumin-TAG and the desired protein.

More specific methods for isolation using the TAG label is provided in the Examples. Some desired proteins may require additions or modifications of the above-described approach as known to one of ordinary skill in the art. The method is scaleable from the laboratory bench to pilot and production facility largely because the techniques applied are well documented in each of these settings.

It is believed that a typical chicken egg produced by a transgenic animal of the present invention will contain at least 0.001 mg, from about 0.001 to 1.0 mg, or from about 0.001 to 100.0 mg of exogenous protein, peptide or polypeptide, in addition to the normal constituents of egg white (or possibly replacing a small fraction of the latter). One of skill in the art will recognize that after biological expression or purification, the desired proteins, fragments thereof and peptides may possess a conformation substantially different than the native conformations of the proteins, fragments thereof and peptides. In this case, it is often necessary to denature and reduce protein and then to cause the protein to re-fold into the preferred conformation. Methods of reducing and denaturing proteins and inducing re-folding are well known to those of skill in the art.

Production of Proteins and Multimeric Proteins in Milk

In addition to methods of producing eggs containing transgenic proteins or peptides, the present invention encompasses methods for the production of milk containing transgenic proteins or peptides. These methods include the administration of a transposon-based vector described above to a mammal, for example through administration into the mammary duct system in order to access epithelial cells. In another embodiment, the transposon-based vector may be administered into arteries supplying the mammary gland. In a further embodiment germline transformed animals may have the transposon-based vector stably incorporated into the mammary epithelial cells. In one embodiment, the transposon-based vector contains a transposase operably-linked to a constitutive promoter and a gene of interest operably-linked to mammary specific promoter. Genes of interest can include, but are not limited to antiviral and antibacterial proteins, immunoglobulins, antibodies and proteins associated with disease and infectious organisms. Milk may be processed for isolation of proteins, multimeric proteins and antibodies using techniques known to one of skill in the art of processing milk and purifying proteins and antibodies from milk.

Administration of the Vaccines Produced with the Present Invention

The proteins, polypeptides, peptides and antibodies made with the present invention for use in vaccines may be administered to animals or humans using techniques known to one of ordinary skill in the art. Clinical practitioners of ordinary skill in the in the veterinary and human medical fields are familiar with methods, routes, volumes and schedules of vaccine administration, as well as various adjuvants that may be optionally employed to engender or amplify an immune response.

The following examples will serve to further illustrate the present invention without, at the same time, however, constituting any limitation thereof. On the contrary, it is to be clearly understood that resort may be had to various embodiments, modifications and equivalents thereof which, after reading the description herein, may suggest themselves to those skilled in the art without departing from the spirit of the invention.

EXAMPLE 1

Preparation of Transposon-Based Vector pTnMod

A vector was designed for inserting a desired coding sequence into the genome of eukaryotic cells, given below as SEQ ID NO:57. The vector of SEQ ID NO:57, termed pTnMod, was constructed and its sequence verified.

This vector employed a cytomegalovirus (CMV) promoter. A modified Kozak sequence (ACCATG) (SEQ ID NO:5) was added to the promoter. The nucleotide in the wobble position in nucleotide triplet codons encoding the first 10 amino acids of transposase was changed to an adenine (A) or thymine (T), which did not alter the amino acid encoded by this codon. Two stop codons were added and a synthetic polyA was used to provide a strong termination sequence. This vector uses a promoter designed to be active soon after entering the cell (without any induction) to increase the likelihood of stable integration. The additional stop codons and synthetic polyA insures proper termination without read through to potential genes downstream.

The first step in constructing this vector was to modify the transposase to have the desired changes. Modifications to the transposase were accomplished with the primers High Efficiency forward primer (Hef) Altered transposase (ATS)-Hef 5' ATCTCGAGACCATGTGTGAACTTGATATTTTACATGATTCTCTTTACC 3'

(SEQ ID NO:91) and Altered transposase- High efficiency reverse primer (Her) 5' GATTGATCATTATCATAATTTCCCCAAAGCGTAACC 3' (SEQ ID NO:92, a reverse complement primer). In the 5' forward primer ATS-Hef, the sequence CTCGAG (SEQ ID NO:93) is the recognition site for the restriction enzyme Xho I, which permits directional cloning of the amplified gene. The sequence ACCATG (SEQ ID NO:5) contains the Kozak sequence and start codon for the transposase and the underlined bases represent changes in the wobble position to an A or T of codons for the first 10 amino acids (without changing the amino acid coded by the codon). Primer ATS-Her (SEQ ID NO:92) contains an additional stop codon TAA in addition to native stop codon TGA and adds a Bcl I restriction site, TGATCA (SEQ ID NO:94), to allow directional cloning. These primers were used in a PCR reaction with pTnLac (p defines plasmid, tn defines transposon, and lac defines the beta fragment of the lactose gene, which contains a multiple cloning site) as the template for the transposase and a FailSafe™ PCR System (which includes enzyme, buffers, dNTP's, $MgCl_2$ and PCR Enhancer; Epicentre Technologies, Madison, WI). Amplified PCR product was electrophoresed on a 1% agarose gel, stained with ethidium bromide, and visualized on an ultraviolet transilluminator. A band corresponding to the expected size was excised from the gel and purified from the agarose using a Zymo Clean Gel Recovery Kit (Zymo Research, Orange, CA). Purified DNA was digested with restriction enzymes Xho I (5') and Bcl I (3') (New England Biolabs, Beverly, MA) according to the manufacturer's protocol. Digested DNA was purified from restriction enzymes using a Zymo DNA Clean and Concentrator kit (Zymo Research).

Plasmid gWhiz (Gene Therapy Systems, San Diego, CA) was digested with restriction enzymes Sal I and BamH I (New England Biolabs), which are compatible with Xho I and Bcl I, but destroy the restriction sites. Digested gWhiz was separated on an agarose gel, the desired band excised and purified as described above. Cutting the vector in this manner facilitated directional cloning of the modified transposase (mATS) between the CMV promoter and synthetic polyA.

To insert the mATS between the CMV promoter and synthetic polyA in gWhiz, a Stratagene T4 Ligase Kit (Stratagene, Inc. La Jolla, CA) was used and the ligation set up according to the manufacturer's protocol. Ligated product was transformed into *E. coli* Top10 competent cells (Invitrogen Life Technologies,

Carlsbad, CA) using chemical transformation according to Invitrogen's protocol. Transformed bacteria were incubated in 1 ml of SOC (GIBCO BRL, CAT# 15544-042) medium for 1 hour at 37° C before being spread to LB (Luria-Bertani media (broth or agar)) plates supplemented with 100 μg/ml ampicillin (LB/amp plates). These plates were incubated overnight at 37° C and resulting colonies picked to LB/amp broth for overnight growth at 37° C. Plasmid DNA was isolated using a modified alkaline lysis protocol (Sambrook et al., 1989), electrophoresed on a 1% agarose gel, and visualized on a U.V. transilluminator after ethidium bromide staining. Colonies producing a plasmid of the expected size (approximately 6.4 kbp) were cultured in at least 250 ml of LB/amp broth and plasmid DNA harvested using a Qiagen Maxi-Prep Kit (column purification) according to the manufacturer's protocol (Qiagen, Inc., Chatsworth, CA). Column purified DNA was used as template for sequencing to verify the changes made in the transposase were the desired changes and no further changes or mutations occurred due to PCR amplification. For sequencing, Perkin-Elmer's Big Dye Sequencing Kit was used. All samples were sent to the Gene Probes and Expression Laboratory (LSU School of Veterinary Medicine) for sequencing on a Perkin-Elmer Model 377 Automated Sequencer.

Once a clone was identified that contained the desired mATS in the correct orientation, primers CMVf-NgoM IV (5' TTGCCGGCATCAGATTGGCTAT (SEQ ID NO:95); underlined bases denote NgoM IV recognition site) and Syn-polyA-BstE II (5' AGAGGTCACCGGGTCAATTCTTCAGCACCTGGTA (SEQ ID NO:96); underlined bases denote BstE II recognition site) were used to PCR amplify the entire CMV promoter, mATS, and synthetic polyA for cloning upstream of the transposon in pTnLac. The PCR was conducted with FailSafe™ as described above, purified using the Zymo Clean and Concentrator kit, the ends digested with NgoM IV and BstE II (New England Biolabs), purified with the Zymo kit again and cloned upstream of the transposon in pTnLac as described below.

Plasmid pTnLac was digested with NgoM IV and BstE II to remove the ptac promoter and transposase and the fragments separated on an agarose gel. The band corresponding to the vector and transposon was excised, purified from the agarose, and dephosphorylated with calf intestinal alkaline phosphatase (New England Biolabs) to prevent self-annealing. The enzyme was removed from the vector using a Zymo DNA Clean and Concentrator-5. The purified vector and CMVp/mATS/polyA

were ligated together using a Stratagene T4 Ligase Kit and transformed into *E. coli* as described above.

Colonies resulting from this transformation were screened (mini-preps) as describe above and clones that were the correct size were verified by DNA sequence analysis as described above. The vector was given the name pTnMod (SEQ ID NO:57) and includes the following components:

Base pairs 1-130 are a remainder of F1(-) on from pBluescriptll sk(-) (Stratagene), corresponding to base pairs 1-130 of pBluescriptll sk(-).

Base pairs 131 - 132 are a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 133 -1777 are the CMV promoter/enhancer taken from vector pGWiz (Gene Therapy Systems), corresponding to bp 229-1873 of pGWiz. The CMV promoter was modified by the addition of an ACC sequence upstream of ATG.

Base pairs 1778-1779 are a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 1780 - 2987 are the coding sequence for the transposase, modified from Tn10 (GenBank accession J01829) by optimizing codons for stability of the transposase mRNA and for the expression of protein. More specifically, in each of the codons for the first ten amino acids of the transposase, G or C was changed to A or T when such a substitution would not alter the amino acid that was encoded.

Base pairs 2988-2993 are two engineered stop codons.

Base pair 2994 is a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 2995 - 3410 are a synthetic polyA sequence taken from the pGWiz vector (Gene Therapy Systems), corresponding to bp 1922-2337 of 10 pGWiz.

Base pairs 3415 - 3718 are non-coding DNA that is residual from vector pNK2859.

Base pairs 3719 - 3761 are non-coding λ DNA that is residual from pNK2859.

Base pairs 3762 - 3831 are the 70 bp of the left insertion sequence recognized by the transposon Tn10.

Base pairs 3832-3837 are a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 3838 - 4527 are the multiple cloning site from pBluescriptll sk(20), corresponding to bp 924-235 of pBluescriptll sk(-). This multiple cloning site may be used to insert any coding sequence of interest into the vector.

Base pairs 4528-4532 are a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 4533 - 4602 are the 70 bp of the right insertion sequence recognized by the transposon Tn10.

Base pairs 4603 - 4644 are non-coding λ DNA that is residual from pNK2859.

Base pairs 4645 - 5488 are non-coding DNA that is residual from pNK2859.

Base pairs 5489 - 7689 are from the pBluescriptll sk(-) base vector - (Stratagene, Inc.), corresponding to bp 761-2961 of pBluescriptll sk(-).

Completing pTnMod is a pBlueScript backbone that contains a colE I origin of replication and an antibiotic resistance marker (ampicillin).

It should be noted that all non-coding DNA sequences described above can be replaced with any other non-coding DNA sequence(s). Missing nucleotide sequences in the above construct represent restriction site remnants.

All plasmid DNA was isolated by standard procedures. Briefly, *Escherichia coli* containing the plasmid was grown in 500 mL aliquots of LB broth (supplemented with an appropriate antibiotic) at 37°C overnight with shaking. Plasmid DNA was recovered from the bacteria using a Qiagen Maxi-Prep kit (Qiagen, Inc., Chatsworth, CA) according to the manufacturer's protocol. Plasmid DNA was resuspended in 500 μL of PCR-grade water and stored at -20°C until used.

EXAMPLE 2

Transposon-Based Vector pTnMCS

Another transposon-based vector was designed for inserting a desired coding sequence into the genome of eukaryotic cells. This vector was termed pTnMCS and its constituents are provided below. The sequence of the pTnMCS vector is provided in SEQ ID NO:56. The pTnMCS vector contains an avian optimized polyA sequence operably-linked to the transposase gene. The avian optimized polyA sequence contains approximately 75 nucleotides that precede the A nucleotide string.

Bp 1 – 130 Remainder of F1 (-) ori of pBluescriptII sk(-) (Stratagene) bp1-130

Bp 133 – 1777 CMV promoter/enhancer taken from vector pGWIZ (Gene Therapy Systems) bp 229-1873

Bp 1783 – 2991 Transposase, from Tn10 (GenBank accession #J01829) bp 108-1316

Bp 2992 – 3344 Non coding DNA from vector pNK2859

Bp 3345 – 3387 Lambda DNA from pNK2859

Bp 3388 – 3457 70 bp of IS10 left from Tn10

Bp 3464 – 3670 Multiple cloning site from pBluescriptII sk(-), thru the XmaI site bp 924-718

Bp 3671 - 3715 Multiple cloning site from pBluescriptII sk(-), from the XmaI site thru the XhoI site. These base pairs are usually lost when cloning into pTnMCS bp 717-673

Bp 3716 – 4153 Multiple cloning site from pBluescriptII sk(-), from the XhoI site bp 672-235

Bp 4159 - 4228 70 bp of IS10 right from Tn10

Bp 4229 - 4270 Lambda DNA from pNK2859

Bp 4271 – 5114 Non-coding DNA from pNK2859

Bp 5115 - 7315 pBluescript sk (-) base vector (Stratagene, Inc.) bp 761-2961.

EXAMPLE 3

Production of Antibody in Egg White

A transposon-based vector containing a CMV promoter/cecropin prepro/antibody heavy chain/cecropin pro/Antibody light chain/conalbumin poly A (SEQ ID NO:97) was injected into the oviduct of quail and chickens. A total of 20 birds were injected (10 chickens and 10 quail) and eggs were harvested from the birds once the eggs were laid. Partially purified egg white protein (EW) was then run on a gel under both reducing and non-reducing conditions. Figure 5 is a picture of the gel. Lanes 1 & 18: molecular weight markers, Lanes 2 and 3: EW #1, non-reduced, reduced, respectively; Lanes 4 and 5: EW #2, non-reduced, reduced, respectively, Lanes 6 and 7: EW #3, non-reduced, reduced, respectively, Lanes 8 and 9: EW #4, non-reduced, reduced, respectively; Lanes 10 and 11: EW #5, non-reduced, reduced, respectively; Lanes 12 and 13: EW #6, non-reduced, reduced, respectively; Lanes 14 and 15: EW #7, non-reduced, reduced, respectively; and Lanes 16 and 17: EW #8 Control, non-reduced, reduced, respectively. Based upon the gel results, the

possibility that the egg white in the treated chicken and quail contains antibody produced by the above-mentioned transposon-based vector cannot be excluded.

EXAMPLE 4

Additional Transposon-Based Vectors for Administration to an Animal

The following example provides a description of various transposon-based vectors of the present invention and several constructs for insertion into the transposon-based vectors of the present invention, all for intraoviduct administration. These examples are not meant to be limiting in any way. The constructs for insertion into a transposon-based vector are provided in a cloning vector pTnMCS or pTnMod, both described above.

pTnMOD (CMV-prepro-HCPro-Lys-CPA) (SEQ ID NO:97)

Bp 1–4090 from vector pTnMod, bp 1 - 4090

Bp 4096–5739 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1864

Bp 5746-5916 Capsite/Prepro taken from GenBank accession # X07404, bp 563–733

Bp 5923-7287 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7288-7302 Pro taken from GenBank accession # X07404, bp 719-733 (includes Lysine)

Bp 7309–7953 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7960–8372 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058

Bp 8374–11973 from cloning vector pTnMod, bp 4091-7690

pTnMCS (CHOVep-prepro-HCPro-CPA) (SEQ ID NO:98)

Bp 1–3715 from vector pTnMCS, bp 1-3715

Bp 3721–4395 Chicken Ovalbumin enhancer taken from GenBank accession # S82527.1, bp 1–675

Bp 4402-5738 Chicken Ovalbumin promoter taken from GenBank accession # J00899-M24999, bp 1-1336

Bp 5745–5915 Capsite/Prepro taken fron GenBank accession # X07404, bp 563–733

Bp 5922-7286 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7287-7298 Pro taken from GenBank accession # X07404, bp 719-730 (does not include Lysine)

Bp 7305–7949 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7956–8363 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058

Bp 8365–11964 from cloning vector pTnMCS, bp 3716-7315

pTnMCS(CHOvep-prepro-HCPro-Lys-CPA) (SEQ ID NO:99)

Bp 1 – 3715 from vector pTnMCS, bp 1-3715

Bp 3721 – 4395 Chicken Ovalbumin enhancer taken from GenBank accession # S82527.1, bp 1–675

Bp 4402 - 5738 Chicken Ovalbumin promoter taken from GenBank accession # J00899-M24999, bp 1-1336

Bp 5745 – 5915 Capsite/Prepro taken fron GenBank accession # X07404, bp 563–733

Bp 5922 - 7286 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7287 - 7301 Pro taken from GenBank accession # X07404, bp 719-733 (includes Lysine)

Bp 7308 – 7952 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7959 – 8366 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058

Bp 8368 – 11967 from cloning vector pTnMCS, bp 3716-7315

pTnMCS (CMV-prepro-HCPro-CPA) (SEQ ID NO:100)

Bp 1 – 3715 from vector pTnMCS, bp 1-3715

Bp 3721 – 5364 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1864

Bp 5371-5541 Capsite/Prepro taken from GenBank accession # X07404, bp 563–733

Bp 5548 - 6912 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 6913 - 6924 Pro taken from GenBank accession # X07404, bp 719-730 (does not Lysine)

Bp 6931 – 7575 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7582 – 7989 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058

Bp 7991 – 11590 from cloning vector pTnMCS, bp 3716-7315

pTnMCS (CMV-prepro-HCPro-Lys-CPA) (SEQ ID NO:101)

Bp 1 – 3715 from vector pTnMCS, bp 1-3715

Bp 3721 – 5364 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1864

Bp 5371-5541 Capsite/Prepro taken from GenBank accession # X07404, bp 563–733

Bp 5548 - 6912 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 6913 - 6927 Pro taken from GenBank accession # X07404, bp 719-733 (includes Lysine)

Bp 6934 – 7578 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7585 – 7992 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058

Bp 7994 – 11593 from cloning vector pTnMCS, bp 3716-7315

pTnMod (CHOvep-prepro-HCPro-CPA) (SEQ ID NO:102)

Bp 1–4090 from vector pTnMod, bp 1-4090

Bp 4096–4770 Chicken Ovalbumin enhancer taken from GenBank accession # S82527.1, bp 1–675

Bp 4777-6113 Chicken Ovalbumin promoter taken from GenBank accession # J00899-M24999, bp 1-1336

Bp 6120-6290 Capsite/Prepro taken from GenBank accession # X07404, bp 563–733

Bp 6297-7661 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7662-7673 Pro taken from GenBank accession # X07404, bp 719-730 (does not include Lysine)

Bp 7680–8324 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 8331–8738 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058

Bp 8740–12339 from cloning vector pTnMod, bp 3716-7315

pTnMod (CHOvep-prepro-HCPro-LYS-CPA) (SEQ ID NO:103)

Bp 1–4090 from vector pTnMod, bp 1-4090

Bp 4096–4770 Chicken Ovalbumin enhancer taken from GenBank accession # S82527.1, bp 1–675

Bp 4777-6113 Chicken Ovalbumin promoter taken from GenBank accession # J00899-M24999, bp 1-1336

Bp 6120-6290 Capsite/Prepro taken from GenBank accession # X07404, bp 563–733

Bp 6297-7661 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7662-7676 Pro taken from GenBank accession # X07404, bp 719-733 (includes Lysine)

Bp 7683–8327 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 8334–8741 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058

Bp 8743–12342 from cloning vector pTnMod, bp 3716-7315

pTnMod (CMV-prepro-HCPro-CPA) (SEQ ID NO:104)

Bp 1–4090 from vector pTnMod, bp 1-4090

Bp 4096–5739 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1864

Bp 5746-5916 Capsite/Prepro taken from GenBank accession # X07404, bp 563–733

Bp 5923-7287 Heavy Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7288-7299 Pro taken from GenBank accession # X07404, bp 719-730 (does not include Lysine)

Bp 7306–7950 Light Chain gene construct taken from antibody RM2 provided by Mark Glassy (Shantha West, Inc)

Bp 7557–7969 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058

Bp 7971–11970 from cloning vector pTnMod, bp 3716-7315

Example 5

Production of a Vaccine in Transgenic Chickens against Brucella Melitensis, Collection of the Vaccine from Eggs and Use of the Vaccine to Confer a Protective Response

Brucella species are considered potential biological warfare agents due to the characteristic protracted debilitating disease caused by these organisms. The antigens of interest are the Group 3 Omps which have been shown to demonstrate a role in protective immunity. Group 3 Omps are comprised of two separate proteins of 25 kDa (Omp25) and 31 kDa (Omp31) (Cloeckaert et al., 1996a).

First, genes encoding *Brucella melitensis* outer membrane proteins (Omp 25 and Omp 31) were cloned for expression into egg albumin. The vectors that are used have been described above and the genes of interest encode for Omp 25 and Omp 31.

Harvest of Brucella Genomic DNA: Genomic DNA from *B. melitensis* is harvested from bacteria grown in 250 ml of *Brucella* Broth (Difco Laboratories, Detroit, MI) for 18 hours at 37° C. The culture is killed by the addition of chloroform for 1 h to prevent accidental infection of lab personnel. After killing, genomic DNA isharvested using a Qiagen Genomic-Tip System (Qiagen, Inc., Valencia, CA) according to the manufacturer's protocol and resuspended in PCR grade water.

PCR Optimization: For PCR optimization, primers specific to Omp25 and Omp31 are used to amplify each gene from genomic DNA of *B. melitensis.* Conditions for PCR with each set of primers are optimized using a FailSafe™ PCR System kit (Epicentre Technologies, Madison, WI), the PCR fragments separated by agarose gel electrophoresis, excised from the gel, and purified from the agarose using

a QIAquick Gel Extraction Kit (Qiagen, Inc). Purified DNA corresponding to Omp25 and Omp31 is stored at –20° C until needed.

Plasmid DNA Isolation: All plasmid DNA is isolated by procedures routinely used in the laboratory of Dr. Cooper. Briefly, *Escherichia coli* containing the desired plasmid is grown in 500 ml aliquots of LB broth (supplemented with the appropriate antibiotic) at 37°C overnight with shaking. Plasmid DNA is recovered from the bacteria using a Qiagen Maxi-Prep kit (Qiagen, Inc.) according to the manufacturer's protocol. Plasmid DNA is resuspended in 500 μL of PCR grade water and stored at ⁻20°C until needed.

The vectors are digested with restriction enzymes. Bands corresponding to the vector are excised from the gel, purified from agarose (see above), and ligated (Fast-Link™ DNA Ligation Kit, Epicentre Technologies) overnight with either DNA encoding Omp25 or Omp31. Ligation mixtures are electroporated into *E. coli* STBL4™ (Life Technologies, Rockville, MD), plated onto LB agar plates supplemented with 200 μg/ml ampicillin, and incubated at 37° C overnight.

Resulting colonies are grown at 37° C overnight in 2 ml of LB broth supplemented with 200 μg/ml ampicillin, plasmid DNA harvested using a QIAprep Spin Miniprep Kit (Qiagen, Inc.), and separated by agarose gel electrophoresis. Bacterial isolates containing the expected insert size are frozen at ⁻80° C. The insert is sequenced to determine orientation and verify the sequence. All sequencing is conducted using a BigDye™ Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems, Foster City, CA) according to the manufacturer's protocol and loaded onto a PE Applied Biosystems 310 Genetic Analyzer. Isolates containing Omp25 (pTnOv25) and Omp31 (pTnOv31) with the correct sequence and orientation are grown as described above for Maxi-Prep DNA isolation. Purified DNA will be resuspended in TE buffer (10 mM tris, 5 mM EDTA, pH 8.0) and stored at ⁻20° C until needed.

Next transgenic chickens are developed to produce Omp25 and Omp31 in the albumin portion of their eggs. The experimental males areobtained from commercial sources at one day of age and maintained in brooders until used. After being placed in a treatment group, males will be housed individually in standard caging in temperature controlled spaces appropriate to the age of the birds. They are watered and fed standard Starter Chick chow *ad lib* and kept initially in a 23:1 h light/dark

cycle, stepping down at approximately weekly intervals to a 15h:8h light/dark cycle, as this regimen optimizes sexual maturity and fertility (Duncan 1999).

For transfection, plasmid DNA is complexed with Superfect™ (Qiagen, Inc) in an approximate 1:3 ratio, following the manufacturer's protocol. The DNA is injected (described below) at a rate of 5 μg per testis in a 50 μL volume [Hank's balanced salt solution (HBSS) will be used to bring the total volume to 50 μl]. At least 20 males are used for controls and receive an injection of HBSS and transfecting reagent. All birds are held until sexual maturity before sperm samples are obtained for PCR.

At age 30-98 days, groups of individual males are subjected to transgene delivery by direct intratesticular injection of DNA. Males are anesthetized with isoflourane via a gas machine and positioned on their side on an animal board with cords tractioning wings and feet to allow access to the testes area. The area is swabbed with 0.5% chlorhexidine and a 2 cm dorsolateral incision made over the testis (which is identical to the procedure for caponization). After exposing the testis, a capillary injection device (described below) is used to inject, in 2 or 3 sites, about 50 μl of the DNA/SuperFect complex (Qiagen) (~5 μg DNA per testis). The injections are calculated to suffuse the DNA throughout the whole testis, allowing as much contact of DNA with spermatogonia as possible. Following injection, the wound is closed in two layers with 4-0 absorbable suture, and the contralateral testis similarly exposed and injected. Following surgery, the bird is returned to its cage to recover. Approximately 50 males are used for this experimental regimen to increase the likelihood of obtaining a male having transgenic sperm.

The injection device consists of a 27 ga hypodermic needle cut off at the hub, bent to the desired angle, and welded onto a standard 100 μl capillary tube. The capillary is filled with the DNA solution and placed into a standard mouth pipette so that the injection rate and volume can be accurately controlled hands-free. After injection, males are allowed to reach sexual maturity in order to obtain sperm for PCR. Males with sperm positive for Omp25 or Omp31 are mated with females to produce transgenic F1s. The F1s are verified for copy number of the transgene. These transgenic F1s are raised to maturity.

Next the expressed proteins are purified from the eggs of the transgenic chickens and subsequently used in vaccine efficacy trials. The purification

is achieved with methods presented above for purification of proteins from eggs. The following is only one approach to purify the protein. The initial objective in the purification of expressed proteins in egg white is to eliminate the major component of egg white, namely, the albumin proteins. Li – Chan, et al., 1986 have described a versatile and simple ion – exchange procedure which largely accomplishes this elimination of the albumin.

The egg white is lightly homogenized to reduce viscosity [shear rate of 3140 sec^{-1} reduces albumin viscosity from 21 to 5 centipoises (Tung et al., 1969]. The albumin solution is filtered through two layers of muslin to remove chalazae. The resulting solution is then passed through an industrial cation – exchanger, IMAC-HP333 (Rohm and Haas, Philadelphia, PA) which has been equilibrated at pH 9 with low ionic strength buffer (phosphate or bicarbonate). The albumin passes through the ion – exchange column leaving some ovomucin and all of the avidin and lysozyme on the column.

This procedure allows cationic proteins/peptides to be separated from the albumin at an early stage in the purification. What is required for making this process a general one is to either purify such proteins which are by their nature, highly positively charged (e.g. cecropin – like lytic peptides or their derivatives) or utilize fusion proteins wherein the fusion adduct contains the requisite cationic charge. Also required is that the magnitude of charge on the expressed protein or fusion protein must exceed considerably that of any possible cationic contaminant.

Once the cationic entities have been captured by the ion – exchanger, elution is initiated by a series of increasing ionic strength, pH 8.0 bicarbonate buffers. The desired product is retained until it is selectively eluted at the highest ionic strength (ammonium bicarbonate, approaching 2M). The water and most of the ammonium bicarbonate are removed by lyophilization. The recovered fusion protein is dissolved in 88% formic acid and subjected to cyanogen bromide cleavage. If the desired component is anionic, the lyophilized residue is dissolved in dilute acetic acid and passed through and anion – exchange resin contained in a jacketed column. The column is washed with dilute hydrochloric acid (0.05 M) and cold (-5°C) ethylene glycol solution pumped through the jacket on the column. By-products from the cleavage reaction are substantially removed by this step. The anionic protein is then lyophilized (using solid KOH in the vacuum system to capture the HCl) to dryness,

dissolved in a minimal amount of water and lyophilized (twice) to remove residual HCl.

If the desired component of the fusion protein is cationic then the cleavage mixture is diluted 5 – fold with acetate buffer and passed onto a sulfoethylcellulose cation exchanger contained in a jacketed column. The column is washed with 0.05 M HCl and then 0.15 M HCl to remove residual cationic materials including degradation products while pumping cold (-5°C) ethylene glycol through the column jacket. The desired product is eluted with 1.0 M HCL adjusted carefully to pH 2.7 by the extremely slow addition of very cold ammonium hydroxide. The eluted protein is lyophilized to dryness using solid KOH within the vacuum system to capture the HCl. The protein is dissolved in a minimum amount of water and lyophilized (twice) to remove residual HCl. Protein quantity will be determined using a Bio-Rad Protein Assay Kit (Bio-Rad Laboratories, Hercules, CA). Purity is determined with Western blots, HPLC, ELISAs and other methods known to one of ordinary skill in the art.

Vaccination and Challenge of Mice: The BALB/c murine model of brucellosis established by Montaraz and Winter (1986) has provided much of the current understanding of immunity to brucellosis. The model is valuable since the murine colonization profile of virulent type strains and vaccine candidates mimics what is often attenuated in the ruminant host (Montaraz and Winter, 1986; Winter, 1990; Schurig et al., 1991; Elzer et al., 1996; Stevens et al., 1997; Edmonds et al., 2000a; Edmonds et al., 2000b). The murine model is also useful due to practical concerns including availability of reagents, cost, and population homogeneity (Montaraz and Winter, 1986; Garcia-Carrillo, 1990).

Bacterial Strains: *Brucella melitensis* strain 16M has been documented to cause chronic infections in mice (Jiminez et al., 1994; Philips et al., 1995). Infectious doses containing approximately 1×10^9 and 1×10^{11} CFU/ml are prepared as described (Elzer et al., 1994c) and snap frozen in liquid nitrogen prior to storage at -80°C. Immediately prior to animal inoculation, samples are thawed and diluted in sterile saline to the appropriate concentration. Doses are verified by serial dilution and plating on Schaedler agar containing 5% bovine blood (SBA) (Difco Laboratories, Detroit, MI).

Female BALB/c mice at approximately 6 to 8 weeks of age are obtained from the Division of Laboratory Animal Medicine at the School of Veterinary Medicine at Louisiana State University (Baton Rouge, LA) and held at least one week prior to use. All animals are housed in a restricted access facility. All mice are used in accordance with the Animal Care and Use Committee at Louisiana State University, and approved protocols will be followed.

Murine Vaccine Efficacy Trials: Mice are divided into groups of ten and are given the following vaccines intraperitoneal (i.p.): Group 1 – saline control; Group 2 – 5 x 10^4 CFU of *B. melitensis* vaccine strain Rev-1 i.v.; Group 3 – Titremax adjuvant; Group 4 – Omp25 plus Titremax adjuvant; Group 5 – Omp31 plus Titremax adjuvant; and Group 6 – Omp25 and Omp31 plus Titremax adjuvant. Mice will be vaccinated with 30 μg *B. melitensis* Omp25 or Omp31. The dose of the Rev-1 strain is based on previously published work in BALB/c mice (Jimenez de Bagues et al., 1994a; Jimenez de Bagues et al., 1994b). Prior to vaccination and 8 weeks post-vaccination, mice were bled and the serum pooled for each group and the Omp25 and Omp31 vaccinated mice are boosted.

Ten weeks following initial vaccination with either Omp25, Omp31, Omp25 plus Omp31, or *B. melitensis* strain Rev. 1, both vaccinated and non-vaccinated mice are i.v. inoculated with the challenge strain 5 x 10^4 CFU of *B. melitensis* strain 16M (Jimenez de Bagues et al., 1994a; Jimenez de Bagues et al., 1994b). Two weeks post-challenge, all mice are euthanized by halothane overdose. At necropsy, the whole spleen is removed by aseptic technique for bacteriological analysis while blood is obtained for serological analysis. Spleens are homogenized with a Sorvall Omni-Mixer (Newton, CT) in sterile phosphate-buffered-saline (PBS) and subsequently serially diluted and plated on SBA. All plates are incubated at 37°C in a 5% CO_2 atmosphere for 3 to 4 days until isolated colonies appear.

Serological procedures: Blood is collected from all the mice before inoculation with *B. melitensis* and at two-week intervals thereafter until they are killed. The sera obtained is stored at -80°C. *Brucella*-specific antibodies are detected by the Card test and Western blot analysis with whole-cell lysates of *B. melitensis* (Elzer et al., 1994a).

Vaccination and challenge of Goats: *Strains*

B. melitensis is of established virulence in pregnant goats (Meador and Deyoe 1986, Enright 1990). The construction and phenotypic characteristics of the isogenic 2308 *htrA* mutant, PHE1, have been described by Elzer et al., (1994a). Infectious doses containing approximately 10^9 or 10^{11} colony forming units (cfu) ml^{-1} in *Brucella* broth (Difco, Detroit, MI, USA) were prepared as described by Elzer et al., (1994a) and snap frozen in a dry ice-ethanol bath before being stored at ⁻80°C. These preparations were thawed just before use and their content of viable brucellae was assessed by serial dilution and plating.

Serological procedures: Blood is collected from all the goats before they are vaccinated, and at two-week intervals thereafter until they are killed. The sera obtained is stored at -80°C. *Brucella*-specific antibodies are detected by the Card test and Western blot analysis with whole-cell lysates of *B abortus* (Elzer et al., 1994a).

Vaccine Efficacy Study: Seventy-two mature, non-pregnant female goats are divided into 6 groups and given 100 μg of the following vaccines subcutaneously (s.c.): Group 1 – saline control; Group 2 – 5 x 10^4 CFU of *B. melitensis* vaccine strain Rev-1 i.v.(this group receives saline first and then Rev-1 2 weeks later); Group 3 – Titremax adjuvant; Group 4 – Omp25 plus Titremax adjuvant; Group 5 – Omp31 plus Titremax adjuvant; and Group 6 – Omp25 and Omp31 plus Titremax adjuvant. Eight weeks later, the 3 Omp groups and adjuvant control group are boosted with a second dose of the appropriate formulation. Twelve other animals are vaccinated SC with 1 x 10^6 CFU of the vaccine *B. melitensis* strain Rev. 1. Previous experiments in goats have shown that this dose of strain Rev. 1 provides protection against challenge with virulent *B. melitensis* strain 16M (Elzer et al., 1998). The remaining 12 animals serve as controls and receive another SC inoculation of the saline diluent. Four weeks after vaccination, a fertile male goat is housed with each group. Following ultrasound confirmation of pregnancy, all animals are conjunctively challenged at approximately 110 days gestation with 1 x 10^7 CFU of virulent *B. melitensis* strain 16M. Following inoculation, animals are monitored daily for delivery. At parturition, the birth status of the kid is recorded and within 24 h, necropsy samples obtained for bacteriologic analysis from the lung and abomasal fluid. Within one month of delivery, dams are

euthanized and samples taken for bacteriologic analysis from the parotid lymph node, prescapular lymph node, supramammary lymph node, internal iliac lymph node, liver, spleen, milk, and uterine fluid. Tissue samples are frozen at -4°C for later analysis; milk and uterine fluid are swabbed the day of collection. Serum samples are obtained prior to vaccination, at challenge, and at necropsy for serological analysis.

Bacteriological procedures: All the tissue specimens are homogenized in sterile PBS (0·002M NaH_2P0_4, 0·008M Na_2HPO_4, 0·15M NaCl, pH 7·2), plated on to Farrell's selective medium (Farrell 1974) and Schaedler agar supplemented with 5% sheep blood. The plates are incubated for 14 days at 37°C in an atmosphere of 5% carbon dioxide. The isolates were identified as *Brucella* species by their Gram-staining characteristics, colonial morphology, urease and oxidase reactions and agglutination with *B. abortus*-specific antibodies in a standard slide agglutination assay. (Elzer et al., 1994a)

Mice Studies: Mice vaccinated with the subcellular components (OMP) in conjunction with the adjuvant Titermax exhibit increased protective immunity (i.e., lower CFU/gm of tissue) as compared to the saline controls. Serological analysis remains negative since the *Brucella* diagnostic tests are dependent on antibody reactions to the O-side chain of the organism. The adjuvant alone does not contribute to the immune response or interfere with the serological analyses. The combination of the Titermax and OMP vaccines results in a predominant TH1 protective immune response as demonstrated by a lower bacterial load in cultured tissues.

Goat Studies: Goats receiving the Titermax and OMP combinations exhibit a protective immune response post-challenge as demonstrated by decreased colonization of the dam/fetus pairs and/or decreased abortions as compared to the saline or Titermax alone controls. *B. melitensis* Rev 1 is the accepted vaccine for caprine brucellosis; it is the "gold standard" for protection in this model. The Titermax/OMP vaccinations results in equivalent or better protection as compared to Rev 1. However, unlike Rev 1, the Titermax/OMP groups remain serologically negative on all diagnostic tests until after challenge when all groups become positive.

All patents, publications and abstracts cited above are incorporated herein by reference in their entirety. It should be understood that the foregoing relates only to preferred embodiments of the present invention and that numerous modifications or alterations may be made therein without departing from the spirit and the scope of the present invention as defined in the following claims.

Appendix A

SEQ ID NO:1 (cecropin pro)
GCG CCA GAG CCG AAA

SEQ ID NO:2 (cecropin pro extended)
GCG CCA GAG CCG AAA TGG AAA GTC TTC AAG

SEQ ID NO:3 (cecropin prepro)
AAT TTC TCA AGG ATA TTT TTC TTC GTG TTC GCT TTG GTT CTG GCT TTG TCA ACA
GTT TCG GCT GCG CCA GAG CCG AAA

SEQ ID NO:4 (cecropin prepro extended)
AAT TTC TCA AGG ATA TTT TTC TTC GTG TTC GCT TTG GTT CTG GCT TTG TCA ACA
GTT TCG GCT GCG CCA GAG CCG AAA TGG AAA GTC TTC AAG

SEQ ID NO:5 (modified Kozak sequence)
ACCATG

SEQ ID NO:6
atg ctg ggc atc tgg acc ctc cta cct ctg gtt ctt acg tct gtt gct aga tta

SEQ ID NO:7
atg att cct gcc aga ttt gcc ggg gtg ctg ctt gct ctg gcc ctc att ttg cca
ggg acc ctt tgt

SEQ ID NO:8
atgg gcagagcaat ggtggccagg ctggggctgg ggctgctgct gctggcactg ctcctaccca
cgcagattta ttcc

SEQ ID NO:9
atgaatctat cgaacatttc tgcggtaaaa gtattaacac tggtggttag cgctgccatc gct

SEQ ID NO:10
atgaccatcc ttttccttac tatggttatc tcatacttca gttgcatgaa agctgccccg
atgaaagaag ctagtgtaag aggacatggc agcttggctt acccaggtct tcggacccac
gggactcttg aaagcctaac tgggcccaat gctggttcaa gaggactgac atcactggcg
gacacttttg aacacgtgat agaggagctt ctagatgaag atcaggacat ccagcccagt
gaggaaaaca aggatgcgga cttgtacaca tcccgagtca tgctgagcag tcaagtgcct
ttggaacccc cactgctctt tctgctcgag gagtacaaaa actacctgga tgctgcaaac
atgtccatga gagtccggcg tcactctgac

SEQ ID NO:11
cgtctttttc tcttatcttt tctcgctttc gctcttttct cgtcggcgat tgctttctcc
gacgacgatc cgttgatccg acaagttgta tcgggaaacg atgacaacca tatgttaaac
gccgagcatc acttttcact tttt

SEQ ID NO:12
atg tccatcttgt tttatgtgat atttcttgca tatcttcgtg gcattcagtc aactaatatg
gatcaaagga gtttgccaga agattcaatg aattctctca ttattaaact cattcgggca
gacatcttga aaaacaagct ttctaagcag gtgatggatg tcaaggaaaa ctatcaaaac
atagtgcaga aagtagagga ccaccaggag atggatggag atgaaaatgt gaaatcagac
ttccagccag ttatttcaat ggatacagac ctcctaaggc agcagagacg ctacaactct
cctcgagttc tcctaagtga caacacacca ctggaaccac caccactgta cctcacagag
gattatgttg gaagttcagt ggtattaaac agaacctctc gaaggaaaag gt

SEQ ID NO:13
atggtgcatc tgactcctga ggagaagtct gccgttactg ccctgtgggg caaggtgaac
gtggatgaag ttggtggtga ggccctgggc aggctgctgg tggtctaccc ttggacccag
aggttctttg agtcctttgg ggatctgtcc actcctgatg ctgttatggg caaccctaag
gtgaaggctc atggcaagaa agtgctcggt gcctttagtg atggcctggc tcacctggac
aacctcaagg gcacctttgc cacactgagt gagctgcact gtgacaagct gcacgtggat
cctgagaact tcaggctcct gggcaacgtg ctggtctgtg tgctggccca tcactttggc
aaagaattca ccccaccagt gcaggctgcc tatcagaaag tggtggctgg tgtggctaat
gccctggccc acaagtatca ctaagctcgc tttcttgctg tccaatttct attaaaggtt

cctttgttcc ctaagtccaa ctactaaact gggggatatt atgaagggcc ttgagcatct
ggattctgcc taataaaaaa catttatttt cattgc

SEQ ID NO:14
gcatggggac ggcgcttctc cagcgcgggg gctgctttct cctgtgcctt tcgctgctgc
tcctgggctg ctgggcggag ctgggcagcggg

SEQ ID NO:15
cgaaacgatt caaaaccтct ttactgccgt tatttgctgg atttttattg ctgttttatt
tggttctggc aggac

SEQ ID NO:16
ggagtctggg ggaggcttag tgcagcctgg agagtccctg aaactctcct gtgaatccaa
tgaatacgaa ttcccttccc atgacatgtc ttgggtccgc aagactccgg agaagaggct
ggagttggtc gcagccatta atagtgatgg tggtagcacc tactatccag acaccatgga
gagacgattc atcatctcca gagacaatac caagaagacc ctgtacctgc aaatgagcag
tctgaggtct gaggacacag ccttgtatta ctgtgcaaga cacacgatga gcaaaagtta
ctgtgagctc aaactaaaac ctcctgcaga gcatccagga ccagcagggg gcgcggagag
acacagagtt gtgaaat

SEQ ID NO:17
acatccattc ttctgtgagt ttcactcgaa gagcagcgtg tcactgcgga caagccagcc
agctcaccat ggctggacct cccagggtac cagacctctg ggaactggcc ctgagcctca
cttcacggac acaggctgcc cgccaaagtg ggtctcagag caacagtgtg tgcattgctc
gtcacatctt cctcttgctt tgcatgactg actacaccca agaagtgtgc ccctgggagg
aaagcatatt tggcaaccag atcataataa aatcagaaat gcagcaaacc tttaaaatat
ccagacttgg

SEQ ID NO:18
tggaagcaag agggagtatg ctaacttcat g

SEQ ID NO:19
atcaattaca agaggg

SEQ ID NO:20
atgaagttmg catactccct cttgcttcca ttggcaggag tcagtgcttc agtkatcaat
tacaagaga

SEQ ID NO:21
aattcttaat taattattgt ggtgtcacaa taacttttc

SEQ ID NO:22
ccccccggat ccatggccgc taaattcgtc gtggttctgg ccgcttgcgt cgccctgagc
cactcggcta tggtgcgccg caagaagaac ggctac

SEQ ID NO:23
ccccccggat ccatgaaact cctggtcgtg ttcgccatgt gcgtgcccgc tgccagcgct

SEQ ID NO:24
cagtgtacgg cggctcgagg cagaagtccg gacgcata

SEQ ID NO:25
gaattcatta tcagcacgga ccagatttct ggatcaggat agaagtctga cgtttacggt
tttcgaagca cagaacgcat tcgttagcgt aagtgtcacc gtcggtaccg caaaccggac
ggtattccag agtgcaaccg ttcagttcgt tgtagcattt agcttcacga cccagagagt
ccatggatcc cccttccgct gtcttctcag ttccaagcat tgcgattttg ttaagcaacg
cactctcgat tcgtagagcc tcgttgcgtt tgtttgcacg aaccatatg

SEQ ID NO:26
ttcacaggca ggtttttgta gagaggggca tgtcatagtc ctcactgtgg

SEQ ID NO:27
aagcttctcg tgaaaaccaa cccaattagt tagtattgca ttctgtgtac tatagtttgg
aatattaaaa atattttaaa atacctccat tttgcttatc cttttagtga agatgatacc
tgcaaaagac atggctaaag ttatgattgt catgttggca atttgttttc ttacaaaatc
ggatgggaaa tctgttaagt aagtactgtt ttgccttgga attggatttt taatgttgac
tttatcattt cgaagtgggg agctaatggg aagtggccct ctctgtttct cttcttccca
ggaagagat

SEQ ID NO:28
atggctacag gctcccggac gtccctgctc ctggcttttg gcctgctctg cctgccctgg
cttcaagagg gcagtgcc

SEQ ID NO:29
atgaggtct ttgctaatct tggtgctttg
cttcctgccc ctggctgctc tgggg

SEQ ID NO:30
atgc acctgagaat ccacgcgaga cggaaccctc ctcgccggcc ggcctggacg cttgggatct
ggtccctttt ctggggatgt atcgtcagct ct

SEQ ID NO:31
atg gccattagtg gagtccctgt gctaggattt ttcatcatag ctgtgctgat gagcgctcag
gaatcatggg ctatcaaaga agaacatgtg atcatccagg ccgagttcta tctgaatcct
gaccaatcag gcgagtttat gtttgac

SEQ ID NO:32
agggggatc cccggagacc ttcgggtagc aactgtcacc ttgatgctgg cgatcctgag
ctcctcactg gctgagggc

SEQ ID NO:33
atg gtgtgtctga ggctccctgg aggctcctgc atggcagttc tgacagtgac
actgatggtg ctgagctccc cactggcttt ggct

SEQ ID NO:34
1 gaacgattta aggagcgaat actactggta aactaatgga agaaatctgc tgcaccactg
61 gatattggga gtgtgtggca tgcatcctca tcatcaggaa actctaaaaa agaaccgagt
121 ggtgctagcc aaacagctgt tgttgagcga attgttagaa catcttctgg agaaggacat
181 catcaccttg gaaatgaggg agctcatcca ggccaaagtg ggcagtttca gccagaatgt
241 ggaactcctc aacttgctgc ctaagagggg tccccaagct tttgatgcct tctgtgaagc
301 actgagggag accaagcaag gccacctgga ggatatgttg ctcaccaccc tttctgggct
361 tcagcatgta ctcccaccgt tgagctgtga ctacgacttg agtctccctt ttccggtgtg

```
421 tgagtcctgt ccctttaca agaagctccg cctgtcgaca gatactgtgg aacactccct
481 agacaataaa gatggtcctg tctgctttca ggtgaagcct tgcactcctg aattttatca
541 aacacacttc cagctggcat atagttgtca gtcttcggcct cgtggcctag cactggtgtt
601 gagcaatgtg cacttcactg gagagaaaga actggaattt cgctctggag gggatgtgga
661 ccacagtact ctagtcaccc tcttcaagct ttgggctat gacgtccatg ttctatgtga
721 ccagactgca caggaaatgc aagagaaact gcagaatttt gcacagttac ctgcacaccg
781 agtcacggac tcctgcatcg tggcactcct ctcgcatggt gtggagggcg ccatctatgg
841 tgtggatggg aaactgctcc agctccaaga ggtttttcag ctctttgaca acgccaactg
901 cccaagccta cagaacaaac caaaaatgtt cttcatccag gcctgccgtg gaggtgctat
961 tggatccctt gggcacctcc ttctgttcac tgctgccacc gcctcttctg ctctatgaga
1021 ctgatcgtgg ggttgaccaa caagatggaa agaaccacgc aggatcccct gggtgcgagg
1081 agagtgatgc cggtaaagaa aagttgccga agatgagact gcccacgcgc tcagacatga
1141 tatgcggcta tgcctgcctc aaagggactg ccgccacgcg gaacaccaaa cgaggttcct
1201 ggtacatcga ggctcttgct caagtgtttt ctgagcgggc ttgtgatatg cacgtggccg
1261 acatgctggt taaggtgaac gcacttatca aggatcgggа aggttatgct cctggcacag
1321 aattccaccg gtgcaaggag atgtctgaat actgcagcac tctgtgccgc cacctctacc
1381 tgttcccagg acaccctccc acatgatgtc acctccccat catccacgcc aagtggaagc
1441 cactggacca caggaggtgt gatagagcct ttgatcttca ggatgcacgg tttctgttct
1501 gcccctcag ggatgtggga atctcccaga cttgtttcct gtgccatca tctctgcctt
1561 tgagtgtggg actccaggcc agctcctttt ctgtgaagcc ctttgcctgt agagccagcc
1621 ttggttggac ctattgccag gaatgtttca gctgcagttg aagagcctga caagtgaagt
1681 tgtaaacaca gtgtggttat ggggagaggg catataaatt ccccatattt gtgttcagtt
1741 ccagctttg tagatggcac tttagtgatt gctttatta cattagttaa gatgtctgag
1801 agaccatctc ctatcttta tttcattcat atcctccgcc ctttttgtcc tagagtgaga
1861 gtttggaagg tgtccaaatt taatgtagac attatctttt ggctctgaag aagcaaacat
1921 gactagagac gcaccttgct gcagtgtcca gaagcggcct gtgcgttccc ttcagtactg
1981 cagcgccacc cagtggaagg acactcttgg ctcgtttggg ctcaaggcac cgcagcctgt
2041 cagccaacat tgccttgcat ttgtacctta ttgatctttg cccatggaag tctcaaagat
2101 ctttcgttgg ttgtttctct gagctttgtt actgaaatga gcctcgtggg gagcatcaga
2161 gaaggccagg aagaatggtg tgtttcccta gactctgtaa ccacctctct gtctttttcc
2221 ttcctgagaa acgtccatct ctctcccttа ctattcccac tttcattcaa tcaacctgca
2281 cttcatatct agatttctag aaaagcttcc tagcttatct ccctgcttca tatctctccc
2341 ttctttacct tcatttcatc ctgttggctg ctgccaccaa atctgtctag aatcctgctt
2401 tacaggatca tgtaaatgct caaagatgta atgtagttct ttgttcctgc tttctctttc
2461 agtattaaac tctcctttga tattatgtgg ctttttattc agtgccatac atgttattgt
```

2521 tttcaaccta gaaaccttta tccctgctta tctgaaactt cccaacttcc ctgttcttta
2581 agacttttt ttttttttt ttttttttg agacagagtc tcgctctgtc gcccaggctg
2641 gagggcagtg gcacgatctc agctcactgc aagctccaac tcccgggttc acgccattct
2701 cctgcctcag ccttccaagt agctgggact acaggtgccc gccaccgtgc ccggctaatt
2761 tttttgtatt tttagtagag acagggtttc accatgttag ccgggatggt cttgatctcc
2821 tgacctcatg atccacccac ctcagcctcc caaagtgttg ggattacagg cgtgagccac
2881 tgcgcccggg caagaccttt ttttaaaaaa aaaaaaaaaa aaacttccat tctttcttcc
2941 tccagtctgt tctcacataa cagagtagtt ttggttttta attttttttg gttgtttgct
3001 gttttttgtt ttttaaggtg agttctcact atgtttctca gactggtctc gaactcctgg
3061 cctcaagcca tcttcccgcc tcagcctctc aaatagctgg gcttacaggc atgagccacc
3121 acacctggcc aggatttggt tgtttaaata taaatctgat caccccccctg cttagaaccc
3181 ttctgctttc tattaccccct catttaaaat gtaaactctt caccttggtt tatgagaact
3241 ggttcttgcc ttccccttga acctcattaa atggtgattt cttgctaagc tccagcccga
3301 gtggtctcct ctcagcttct aattttgtgc tctttcctgc ccttttcctg ggccttctca
3361 gctctccacc cccaccactc ttgactcagg tggtgtcctt cttcctcaag tcttgacaat
3421 tcccgggccc ttcagtccct gagcagtcta cttctgtgtc tgtcaccaca tcttgtcttt
3481 tcccctcatt gcatttattg cagtttatat atatgctact tttacttgtt catttctgtc
3541 tcccctacca ggctgtaaat gagggcagaa accttgtttg ttttattcac catcatgtac
3601 caagtgcttg gcacatagtg ggccttcatt aaatgtttgt tgaataaaag agggaagaag
3661 gcaagccaac cttagctaca atcctacctt ttgataaaat gttccttttg acaatatata
3721 cggattatta tttgtacttt gtttttccat gtgttttgct tttatccact ggcatttttа
3781 gctccttgaa gacatatcat gtgtgagata acttccttca catctcccat ggtcoctagc
3841 aaaatgctag gcctgtagta gtcaaggtgc tcaataaata tttgtttggg tggtttgtga
3901 gccttgctgc caagtcctgc ctttgggtcg acatagtatg gaagtatttg agagagagaa
3961 cctttccact cccactgcca ggattttgta ttgccatcgg gtgccaaata aatgctcata
4021 tttattaaaa aaaaaaaaaa aaaaa

SEQ ID NO:35
1 tccagatcat ctgtcctcac caccaaggcc atggtgtctt cagcgactat ctgcagtttg
61 ctactcctca gcatgctctg gatggacatg gccatggcag gttccagctt cttgagccca
121 gagcaccaga aagcccagca gagaaaggaa tccaagaagc caccagctaa actgcagcca
181 cgagctctgg aaggctggct ccacccagag gacagaggac aagcagaaga ggcagaggag
241 gagctggaaa tcaggttcaa tgctcccttc gatgttggca tcaagctgtc aggagctcag
301 taccagcagc atggccgggc cctgggaaag tttcttcagg atatcctctg ggaagaggtc
361 aaagaggcgc cagctaacaa gtaaccactg acaggactgg tccctgtact ttcctcctaa

421 gcaagaactc acatccagct tctgcctcct ctgcaactcc cagcactctc ctgctgactt
481 acaaataaat gttcaagctg t

SEQ ID NO:36
atgaagctgc ttgcaatggt tgcactgctg gtcaccatct gtagcctaga aggagctttg
gttcggagac

SEQ ID NO:37
atgaagctgc ttgcaatggt tgcactgctg gtcaccatct gtagcctaga aggagctttg
gttcggagac

SEQ ID NO:38
atggcctt gccaacggct cgacccctgt tggggtcctg tgggaccccc gccctcggca
gcctcctgtt cctgctcttc agcctcggat gggtgcagcc c

SEQ ID NO:39
atgccgcgcc tgttctccta cctcctaggt gtctggctgc tcctgagcca acttcccaga
gaaatcccag gc

SEQ ID NO:40
atgacagc atcacttgtc gttttaccat cgctttggtt aatattaatt atttttactg
caccctatac tcactgt

SEQ ID NO:41
atgaaagtc ctgctttgtg acctgctgct gctcagtctc ttctccagtg tgttcagcag
ttgtcagagg gactgtctca catgccagga gaagctccac ccagccctgg acagcttcga
cctggaggtg tgcatcctcg agtgcgaaga gaaggtcttc cccagccccc tctggactcc
atgcaccaag gtcatggcca ggagctcttg gcagctcagc cctgccgccc cagagcatgt
ggcggctgct ctctaccagc cgagagcttc ggagatgcag catctgcggc gaatgccccg
agtccggagc ttgttccagg agcaggaaga gcccgagcct ggcatggagg aggctggtga
gatggagcag aagcagctgc agaagagat

SEQ ID NO:42
aattcatgaa gtgggttact ttcatctctt tgttgttctt gttctcttct gcttactcta
gaggtgtttt cagacg

SEQ ID NO:43
atgaagtggg ttactttcat ctctttgttg ttcttgttct cttctgctta ctctagaggt
gttttcagac gc

SEQ ID NO:44
1 gaattctcaa tggcaaaggc aagtgtacat tataaatagc aaaacagctg gcttggacca
61 tgttgccggc cagtcaccca gttgagggat ttgaatgaca tcataaccct caagagggta
121 ttgctagcca gctggtgtta tttagaatac acaaaaacca gagaaagaaa acacactctg
181 gcacacagac tccctctgtc atacacacac acacacacac acacacacac acacacacac
241 agaggtttga gttatatgga aaattcaaac aacaggaaaa ttgtttgccc cccaggtacc
301 cttctcccag agtggtgggg tggggagggg acagtgacag gcagcctagt agaagaataa
361 agaaaaatgt tctatttcag ttgggtttta cagctcggca tagtctttgc ctcatcgcag
421 gagaaaaagt atgagacagt gccctaaagg gaccaatcca atgctgcctg cccctccata
481 cgttctagga aatgagatca cacccctcac ttggcaactg ggacaagggg tcacccgagt
541 gctgtcttcc aatctacttt accccagtca cttcagggtt aaaattgtag agtttgctgg
601 agagggtctt atcgtccttt ctttcttttt ttgttttaaa taatgcattt gctctagaat
661 ctaaaattgc tctcccatcc cccatattcc tttaatactg gtaaggtgta ttagcagacg
721 tttgtgtctt catgcccagc agaaagttaa tcagaaaaca gatccttatt ttctatggca
781 gcataagtat tttaatgtct gcgaaccctg tcagtaacac acattctttt aagggaaaaa
841 aatgcttctg tgctctagtt ttaaaatgca aaggtatgat gttatttgtc accatgccca
901 aaaaagtcct tactcaataa ctttgccaga agagggagag agagagaagg caaatgttcc
961 cccagctgtt tcctgtctac agtgtctgtg ttttgtagat aaatgtgagg attttgtgta
1021 aatccctctt ctgtttgcta aatctcactg tcactgctaa attcagagca gatagagcct
1081 gcgcaatgga ataaagtcct caaaattgaa atgtgacatt gctctcaaca tctcccatct
1141 ctctggattt ctttttgctt cattattcct gctaaccaat tcattttcag actttgtact
1201 tcagaagcaa tgggaaaaat cagcagtctt ccaacccaat tatttaagtg ctgcttttgt
1261 gatttcttga aggtaaatat ttcttactct ttgaagtcat tggggaattc ggatcccact
1321 gtaataatag catctttcat ttccgtagta aacgtttcta gatattttgt ctcaattcat
1381 tgaaatagga acccataaag aaaggggttc agggaggact cctccaaaga tccacagtag
1441 ccaggggaat aaacacaggt tgttggatgc cgagacacgc tccatccaca actccctgct

1501 gggttctcat gtactctatt ggcttctgtg ctgggtagtc ctgattaatg acagtcgtgg
1561 aatcgtggga gtcaatgcac ttctgtccca ccccactccc cttgcaagga tcaaggagga
1621 aacctgaacc tcccctctgtt tcttgggcag gtgaagatgc acaccatgtc ctcctcgcat
1681 ctcttctacc tggcgctgtg cctgctcacc ttcaccagct ctgccacggc t

SEQ ID NO:45
atgaag ccaattcaaa aactcctagc tggccttatt ctactgactt cgtgcgtgga
aggctgctcc agccagcact ggtcctatgg actgcgccct ggaggaaaga gagatgccga
aaattt

SEQ ID NO:46
atgagatt tccttcaatt tttactgcag ttttattcgc agcatcctcc gcattagctg
ctccagtcaa cactacaaca gaagatgaaa cggcacaaat tccggctgaa gctgtcatcg
gttacttaga tttagaaggg gatttcgatg ttgctgtttt gccattttcc aacagcacaa
ataacgggtt attgtttata aatactacta ttgccagcat tgctgctaaa gaagaagggg
tatctttgga taaaagagag gctgaagctt

SEQ ID NO:47
atgaa gtgggtaacc tttatttccc ttctttttct ctttagctcg gcttattcca ggggtgtgtt
tcgtcgaga

SEQ ID NO:48
1 atgaagatgg tctcctcctc gcgcctccgc tgcctcctcg tgctcctgct gtccctgacc
61 gcctccatca gctgctcctt cgccggacag agagactcca aactccgcct gctgctgcac
121 cggtacccgc tgcagggctc caaacaggac atgactcgct ccgccttggc cgagctgctc
181 ctgtcggacc tcctgcaggg ggagaacgag gctctggagg aggagaactt ccctctggcc
241 gaaggaggac ccgaggacgc ccacgccgac ctagagcggg ccgccagcgg ggggcctctg
301 ctcgcccccc gggagagaaa ggccggctgc aagaacttct tctggaaaac cttcacctcc
361 tgctga

SEQ ID NO:49
atggccgggc gaggggcag cgcgctgctg gctctgtgcg gggcactggc tgcctgcggg
tggctcctgg gcgccgaagc ccaggagccc ggggcgcccg cggcgggcat gaggcggcgc
cggcggctgc agcaagagga cggcatctcc ttcgagtacc accgctaccc cgagctgcgc

gaggcgctcg tgtccgtgtg gctgcagtgc accgccatca gcaggattta cacggtgggg
cgcagcttcg agggccggga gctcctggtc atcgagctgt ccgacaaccc tggcgtccat
gagcctggtg agcctgaatt taaatacatt gggaatatgc atgggaatga ggctgttgga
cgagaactgc tcatttttctt ggcccagtac ctatgcaacg aataccagaa ggggaacgag
acaattgtca acctgatcca cagtacccgc attcacatca tgccttccct gaacccagat
ggctttgaga aggcagcgtc tcagcctggt gaactcaagg actggtttgt gggtcgaagc
aatgcccagg gaatagatct gaaccggaac tttccagacc tggataggat agtgtacgtg
aatgagaaag aaggtggtcc aaataatcat ctgttgaaaa atatgaagaa aattgtggat
caaaacacaa agcttgctcc tgagaccaag gctgtcattc attggattat ggatattcct
tttgtgcttt ctgccaatct ccatggagga gaccttgtgg ccaattatcc atatgatgag
acgcggagtg gtagtgctca cgaatacagc tcctccccag atgacgccat tttccaaagc
ttggcccggg catactcttc tttcaacccg gccatgtctg accccaatcg gccaccatgt
cgcaagaatg atgatgacag cagctttgta gatggaacca ccaacggtgg tgcttggtac
agcgtacctg gagggatgca agacttcaat taccttagca gcaactgttt tgagatcacc
gtggagctta gctgtgagaa gttcccacct gaagagactc tgaagaccta ctgggaggat
aacaaaaact ccctcattag ctaccttgag cagatacacc gaggagttaa aggatttgtc
cgagaccttc aaggtaaccc aattgcgaat gccaccatct ccgtggaagg aatagaccac
gatgttacat ccgcaaagga tggtgattac tggagattgc ttatacctgg aaactataaa
cttacagcct cagctccagg ctatctggca ataacaaaga aagtggcagt tccttacagc
cctgctgctg gggttgattt tgaactggag tcattttctg aaaggaaaga agaggagaag
gaagaattga tggaatggtg gaaaatgatg tcagaaactt taaatttt

SEQ ID NO:50
atggctctct cactcttcac tgttggacaa ttaattttct tattttggac actcagaatc
actgaagcc

SEQ ID NO:51
atgaacaaa ctagcaattc tcgctatcat cgctatggta cttttcagcg caaacgcctt
cagactccaa agcagattga gatcaaatat ggaagcttct gccagag

SEQ ID NO:52
a tggtcagtgt gtgcaggctc ttgctggttg ctgccttgct gctgtgtttg caagcacagc
tgtctttctc tcagcactgg tctcatggct ggtaccctgg aggaaagaga gaaatcgact
cctacagctc accagagata tctggggaga ttaaactgtg tgaagcggga gaatgcagct
atctcaggcc actgaggacc aacatcctaa agagcatcct gattgacacc cttgcaagga

aattccaaaa gaggaaatga

SEQ ID NO:53
1 tgtgttttgt agataaatgt gaggattttc tctaaatccc tcttctgctt gctaaatctc
61 actgtcgctg ctaaattcag agcagataga gcctgcgcaa tcgaaataaa gtcctcaaaa
121 ttgaaatgtg actttgctct aacatctccc atctctctgg atttcttttt gcctcattat
181 tcctgcccac caattcattt ccagactttg tacttcagaa gcgatgggga aaatcagcag
241 tcttccaact caattattta agatctgcct ctgtgacttc ttgaagataa agatacacat
301 catg

SEQ ID NO:54
atgtc aggcccgagg acgtgcttct gtctaccgtc ggctcttgta ctagtactgc tgagtctcag
cacttcggca ctaggg

SEQ ID NO:55
mspaaqlaka aarstcmtrl psgirvatap snshfaavgv yvdagpiyet sidrgvshfv
sslafksthg atesqvlktm aglggnlfct atresilyqg svlhhdlprt vqlladttlr
palteeeiae rratiafeae dlhsrpdafi gemmhavafg grglgnsifc epqrarnmts
dtireyfaty lhpsrmvvag tgvahaelvd lvskafvpss trapssvths dietayvggs
hqlvipkppp thpnyeqtlt hvqvafpvpp fthpdmfpvs tlqvlmgggg afsaggpgkg
mysrlytnvl nryrwmesca afqhayssts lfgisascvp sfnphlcnvl agefvhmarn
lsdeevarak nqlkssllmn lesqvitved igrqvlaqnq rleplelvnn isavtrddlv
rvaealvakp ptmvavgedl tkltdiketl aafnasgeal qpvgsagsfg rvtm

SEQ ID NO:56 (pTnMCS)
1 ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga
61 ccgctacact tgccagcgcc ctagcgcccg ctcctttcgc tttcttccct tcctttctcg
121 ccacgttcgc cggcatcaga ttggctattg gccattgcat acgttgtatc catatcataa
181 tatgtacatt tatattggct catgtccaac attaccgcca tgttgacatt gattattgac
241 tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg
301 cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt
361 gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca
421 atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc
481 aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta
541 catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac
601 catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg

661 atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg
721 ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt
781 acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg
841 ccatccacgc tgttttgacc tccatagaag acaccgggac cgatccagcc tccgcggccg
901 ggaacggtgc attggaacgc ggattccccg tgccaagagt gacgtaagta ccgcctatag
961 actctatagg cacacccctt tggctcttat gcatgctata ctgttttgg cttgggcct
1021 atacacccc gcttccttat gctataggtg atggtatagc ttagcctata ggtgtgggtt
1081 attgaccatt attgaccact cccctattgg tgacgatact ttccattact aatccataac
1141 atggctcttt gccacaacta tctctattgg ctatatgcca atacctgtc cttcagagac
1201 tgacacggac tctgtatttt tacaggatgg ggtcccattt attatttaca aattcacata
1261 tacaacaacg ccgtcccccg tgcccgcagt ttttattaaa catagcgtgg gatcttccacg
1321 cgaatctcgg gtacgtgttc cggacatggg ctcttctccg gtagcggcgg agcttccaca
1381 tccgagccct ggtcccatgc ctccagcggc tcatggtcgc tcggcagctc cttgctccta
1441 acagtggagg ccagacttag gcacagcaca atgcccacca ccaccagtgt gccgcacaag
1501 gccgtggcgg taggtatgt gtctgaaaat gagcgtggag attgggctcg cacggctgac
1561 gcagatggaa gacttaaggc agcggcagaa gaagatgcag gcagctgagt tgttgtattc
1621 tgataagagt cagaggtaac tcccgttgcg gtgctgttaa cggtggaggg cagtgtagtc
1681 tgagcagtac tcgttgctgc cgcgcgcgcc accagacata atagctgaca gactaacaga
1741 ctgttccttt ccatgggtct tttctgcagt caccgtcgga ccatgtgcga actcgatatt
1801 ttacacgact ctctttacca attctgcccc gaattacact taaaacgact caacagctta
1861 acgttggctt gccacgcatt acttgactgt aaaactctca ctcttaccga acttggccgt
1921 aacctgccaa ccaaagcgag aacaaaacat aacatcaaac gaatcgaccg attgttaggt
1981 aatcgtcacc tccacaaaga gcgactcgct gtataccgtt ggcatgctag ctttatcttgt
2041 tcgggcaata cgatgcccat tgtacttgtt gactggtctg atattcgtga gcaaaaacga
2101 cttatggtat tgcgagcttc agtcgcacta cacggtcgtt ctgttactct ttatgagaaa
2161 gcgttcccgc ttcagagca atgttcaaag aaagctcatg accaatttct agccgacctt
2221 gcgagcattc taccgagtaa caccacaccg ctcattgtca gtgatgctgg ctttaaagtg
2281 ccatggtata aatccgttga gaagctgggt tggtactggt taagtcgagt aagaggaaaa
2341 gtacaatatg cagacctagg agcggaaaac tggaaaccta tcagcaactt acatgatatg
2401 tcatctagtc actcaaagac tttaggctat aagaggctga ctaaaagcaa tccaatctca
2461 tgccaaattc tattgtataa atctcgctct aaaggccgaa aaaatcagcg ctcgacacgg
2521 actcattgtc accacccgtc acctaaaatc tactcagcgt cggcaaagga gccatgggtt
2581 ctagcaacta acttacctgt tgaaattcga acacccaaac aacttgttaa tatcttattcg
2641 aagcgaatgc agattgaaga aaccttccga gacttgaaaa gtcctgccta cggactaggc
2701 ctacgccata gccgaacgag cagctcagag cgttttgata tcatgctgct aatcgccctg
2761 atgcttcaac taacatgttg gcttgcgggc gttcatgctc agaaacaagg ttgggacaag
2821 cacttccagg ctaacacagt cagaaatcga aacgtactct caacagttcg cttaggcatg
2881 gaagttttgc ggcattctgg ctacacaata acaagggaag acttactcgt ggctgcaacc
2941 ctactagctc aaaatttatt cacacatggt tacgctttgg ggaaattatg aggggatcgc

3001 tctagagcga tccgggatct cgggaaaagc gttggtgacc aaaggtgcct tttatcatca
3061 ctttaaaaat aaaaaacaat tactcagtgc ctgttataag cagcaattaa ttatgattga
3121 tgcctacatc acaacaaaaa ctgatttaac aaatggttgg tctgccttag aaagtatatt
3181 tgaacattat cttgattata ttattgataa taataaaaac cttatcccta tccaagaagt
3241 gatgcctatc attggttgga atgaacttga aaaaaattag ccttgaatac attactggta
3301 aggtaaacgc cattgtcagc aaattgatcc aagagaacca acttaaagct ttcctgacgg
3361 aatgttaatt ctcgttgacc ctgagcactg atgaatcccc taatgatttt ggtaaaaatc
3421 attaagttaa ggtggataca catcttgtca tatgatcccg gtaatgtgag ttagctcact
3481 cattaggcac cccaggcttt acactttatg cttccggctc gtatgttgtg tggaattgtg
3541 agcggataac aatttcacac aggaaacagc tatgaccatg attacgccaa gcgcgcaatt
3601 aaccctcact aaagggaaca aaagctggag ctccaccgcg gtggcggccg ctctagaact
3661 agtggatccc ccgggctgca ggaattcgat atcaagctta tcgataccgc tgacctcgag
3721 ggggggcccg gtacccaatt cgccctatag tgagtcgtat tacgcgcgct cactggccgt
3781 cgttttacaa cgtcgtgact gggaaaaccc tggcgttacc caacttaatc gccttgcagc
3841 acatcccccc ttcgccagct ggcgtaatag cgaagaggcc cgcaccgatc gcccttccca
3901 acagttgcgc agcctgaatg gcgaatggaa attgtaagcg ttaatatttt gttaaaattc
3961 gcgttaaatt ttgttaaatc agctcatttt ttaaccaatt aggccgaaat cggcaaaatc
4021 ccttataaat caaaagaata gaccgagata gggttgagtg ttgttccagt ttggaacaag
4081 agtccactat taaagaacgt ggactccaac gtcaaagggc gaaaaaccgt ctatcagggc
4141 gatggcccac tatccggga tcatatgaca agatgtgtat ccaccttaac ttaatgattt
4201 ttaccaaaat cattagggga ttcatcagtg ctcagggtca acgagaatta acattccgtc
4261 aggaaagctt atgatgatga tgtgcttaaa aacttactca atggctggtt atgcatatcg
4321 caatacatgc gaaaaacta aaagagcttg ccgataaaaa aggcaattt attgctattt
4381 accgcggctt tttattgagc ttgaaagata aataaaatag ataggtttta tttgaagcta
4441 aatctttttt atcgtaaaaa atgccctctt gggttatcaa gagggtcatt atatttcgcg
4501 gaataacatc atttggtgac gaaataacta agcacttgtc tcctgtttac tcccctgagc
4561 ttgagggggtt aacatgaagg tcatcgatag caggataata atacagtaaa acgctaaacc
4621 aataatccaa atccagccat cccaaattgg tagtgaatga ttataaataa cagcaaacag
4681 taatgggcca ataacaccgg ttgcattggt aaggctcacc aataatccct gtaaagcacc
4741 ttgctgatga ctctttgttt ggatagacat cactccctgt aatgcaggta aagcgatccc
4801 accaccagcc aataaatta aaacagggaa aactaaccaa ccttcagata taaacgctaa
4861 aaaggcaaat gcactactat ctgcaataaa tccgagcagt actgccgttt tttcgcccat
4921 ttagtggcta ttcttcctgc cacaaaggct tggaatactg agtgtaaaag accaagaccc
4981 gtaatgaaaa gccaaccatc atgctattca tcatcacgat ttctgtaata gcaccacacc
5041 gtgctggatt ggctatcaat gcgctgaaat aataatcaac aaatggcatc gttaaataag
5101 tgatgtatac cgatcagctt ttgttccctt tagtgagggt taattgcgcg cttggcgtaa
5161 tcatggtcat agctgtttcc tgtgtgaaat tgttatccgc tcacaattcc acacaacata
5221 cgagccggaa gcataaagtg taaagcctgg ggtgcctaat gagtgagcta actcacatta
5281 attgcgttgc gctcactgcc cgctttccag tcgggaaacc tgtcgtgcca gctgcattaa

```
5341 tgaatcggcc aacgcgcggg gagaggcggt ttgcgtattg ggcgctcttc cgcttcctcg
5401 ctcactgact cgctgcgctc ggtcgttcgg ctgcggcgag cggtatcagc tcactcaaag
5461 gcggtaatac ggttatccac agaatcaggg gataacgcag gaaagaacat gtgagcaaaa
5521 ggccagcaaa aggccaggaa ccgtaaaaag gccgcgttgc tggcgttttt ccataggctc
5581 cgcccccctg acgagcatca caaaaatcga cgctcaagtc agaggtggcg aaacccgaca
5641 ggactataaa gataccaggc gtttcccccc ggaagctccc tcgtgcgctc tcctgttccg
5701 accctgccgc ttaccggata cctgtccgcc tttctccctt cgggaagcgt ggcgctttct
5761 catagctcac gctgtaggta tctcagttcg gtgtaggtcg ttcgctccaa gctgggctgt
5821 gtgcacgaac cccccgttca gcccgaccgc tgcgccttat ccggtaacta tcgtcttgag
5881 tccaacccgg taagacacga cttatcgcca ctggcagcag ccactggtaa caggattagc
5941 agagcgaggt atgtaggcgg tgctacagag ttcttgaagt ggtggcctaa ctacggctac
6001 actagaagga cagtatttgg tatctgcgct ctgctgaagc cagttacctt cggaaaaaga
6061 gttggtagct cttgatccgg caaacaaacc accgctggta gcggtggttt ttttgtttgc
6121 aagcagcaga ttacgcgcag aaaaaaagga tctcaagaag atcctttgat cttttctacg
6181 gggtctgacg ctcagtggaa cgaaaactca cgttaaggga ttttggtcat gagattatca
6241 aaaaggatct tcacctagat ccttttaaat taaaaatgaa gttttaaatc aatctaaagt
6301 atatatgagt aaacttggtc tgacagttac caatgcttaa tcagtgaggc acctatctca
6361 gcgatctgtc tatttcgttc atccatagtt gcctgactcc ccgtcgtgta gataactacg
6421 atacgggagg gcttaccatc tggccccagt gctgcaatga taccgcgaga cccacgctca
6481 ccggctccag atttatcagc aataaaccag ccagccggaa gggccgagcg cagaagtggt
6541 cctgcaactt tatccgcctc catccagtct attaattgtt gccgggaagc tagagtaagt
6601 agttcgccag ttaatagttt gcgcaacgtt gttgccattg ctacaggcat cgtggtgtca
6661 cgctcgtcgt ttggtatggc ttcattcagc tccggttccc aacgatcaag gcgagttaca
6721 tgatccccca tgttgtgcaa aaaagcggtt agctccttcg gtcctccgat cgttgtcaga
6781 agtaagttgg ccgcagtgtt atcactcatg gttatggcag cactgcataa ttctcttact
6841 gtcatgccat ccgtaagatg cttttctgtg actggtgagt actcaaccaa gtcattctga
6901 gaatagtgta tgcggcgacc gagttgctct tgcccggcgt caatacggga taataccgcg
6961 ccacatagca gaactttaaa agtgctcatc attggaaaac gttcttcggg gcgaaaactc
7021 tcaaggatct taccgctgtt gagatccagt tcgatgtaac ccactcgtgc acccaactga
7081 tcttcagcat cttttacttt caccagcgtt tctgggtgag caaaaacagg aaggcaaaat
7141 gccgcaaaaa agggaataag ggcgacacgg aaatgttgaa tactcatact cttcctttt
7201 caatattatt gaagcattta tcagggttat tgtctcatga gcggatacat atttgaatgt
7261 atttagaaaa ataaacaaat aggggttccg cgcacatttc cccgaaaagt gccac
```

SEQ ID NO:57 (pTnMod)

```
CTGACGCGCC CTGTAGCGGC GCATTAAGCG CGGCGGGTGT GGTGGTTACG 50
```

```
CGCAGCGTGA CCGCTACACT TGCCAGCGCC CTAGCGCCCG CTCCTTTCGC 100
TTTCTTCCCT TCCTTTCTCG CCACGTTCGC CGGCATCAGA TTGGCTATTG 150
GCCATTGCAT ACGTTGTATC CATATCATAA TATGTACATT TATATTGGCT 200
CATGTCCAAC ATTACCGCCA TGTTGACATT GATTATTGAC TAGTTATTAA 250
TAGTAATCAA TTACGGGGTC ATTAGTTCAT AGCCCATATA TGGAGTTCCG 300
CGTTACATAA CTTACGGTAA ATGGCCCGCC TGGCTGACCG CCCAACGACC 350
CCCGCCCATT GACGTCAATA ATGACGTATG TTCCCATAGT AACGCCAATA 400
GGGACTTTCC ATTGACGTCA ATGGGTGGAG TATTTACGGT AAACTGCCCA 450
CTTGGCAGTA CATCAAGTGT ATCATATGCC AAGTACGCCC CCTATTGACG 500
TCAATGACGG TAAATGGCCC GCCTGGCATT ATGCCCAGTA CATGACCTTA 550
TGGGACTTTC CTACTTGGCA GTACATCTAC GTATTAGTCA TCGCTATTAC 600
CATGGTGATG CGGTTTTGGC AGTACATCAA TGGGCGTGGA TAGCGGTTTG 650
ACTCACGGGG ATTTCCAAGT CTCCACCCCA TTGACGTCAA TGGGAGTTTG 700
TTTTGGCACC AAAATCAACG GGACTTTCCA AAATGTCGTA ACAACTCCGC 750
CCCATTGACG CAAATGGGCG GTAGGCGTGT ACGGTGGGAG GTCTATATAA 800
GCAGAGCTCG TTTAGTGAAC CGTCAGATCG CCTGGAGACG CCATCCACGC 850
TGTTTTGACC TCCATAGAAG ACACCGGGAC CGATCCAGCC TCCGCGGCCG 900
GGAACGGTGC ATTGGAACGC GGATTCCCCG TGCCAAGAGT GACGTAAGTA 950
CCGCCTATAG ACTCTATAGG CACACCCCTT TGGCTCTTAT GCATGCTATA 1000
CTGTTTTTGG CTTGGGGCCT ATACACCCCC GCTTCCTTAT GCTATAGGTG 1050
ATGGTATAGC TTAGCCTATA GGTGTGGGTT ATTGACCATT ATTGACCACT 1100
CCCCTATTGG TGACGATACT TTCCATTACT AATCCATAAC ATGGCTCTTT 1150
GCCACAACTA TCTCTATTGG CTATATGCCA ATACTCTGTC CTTCAGAGAC 1200
TGACACGGAC TCTGTATTTT TACAGGATGG GGTCCCATTT ATTATTTACA 1250
AATTCACATA TACAACAACG CCGTCCCCCG TGCCCGCAGT TTTTATTAAA 1300
CATAGCGTGG GATCTCCACG CGAATCTCGG GTACGTGTTC CGGACATGGG 1350
CTCTTCTCCG GTAGCGGCGG AGCTTCCACA TCCGAGCCCT GGTCCCATGC 1400
CTCCAGCGGC TCATGGTCGC TCGGCAGCTC CTTGCTCCTA ACAGTGGAGG 1450
CCAGACTTAG GCACAGCACA ATGCCCACCA CCACCAGTGT GCCGCACAAG 1500
GCCGTGGCGG TAGGGTATGT GTCTGAAAAT GAGCGTGGAG ATTGGGCTCG 1550
CACGGCTGAC GCAGATGGAA GACTTAAGGC AGCGGCAGAA GAAGATGCAG 1600
GCAGCTGAGT TGTTGTATTC TGATAAGAGT CAGAGGTAAC TCCCGTTGCG 1650
GTGCTGTTAA CGGTGGAGGG CAGTGTAGTC TGAGCAGTAC TCGTTGCTGC 1700
CGCGCGCGCC ACCAGACATA ATAGCTGACA GACTAACAGA CTGTTCCTTT 1750
CCATGGGTCT TTTCTGCAGT CACCGTCGGA CCATGTGTGA ACTTGATATT 1800
```

```
TTACATGATT CTCTTTACCA ATTCTGCCCC GAATTACACT TAAAACGACT 1850
CAACAGCTTA ACGTTGGCTT GCCACGCATT ACTTGACTGT AAAACTCTCA 1900
CTCTTACCGA ACTTGGCCGT AACCTGCCAA CCAAAGCGAG AACAAAACAT 1950
AACATCAAAC GAATCGACCG ATTGTTAGGT AATCGTCACC TCCACAAAGA 2000
GCGACTCGCT GTATACCGTT GGCATGCTAG CTTTATCTGT TCGGGAATAC 2050
GATGCCCATT GTACTTGTTG ACTGGTCTGA TATTCGTGAG CAAAAACGAC 2100
TTATGGTATT GCGAGCTTCA GTCGCACTAC ACGGTCGTTC TGTTACTCTT 2150
TATGAGAAAG CGTTCCCGCT TTCAGAGCAA TGTTCAAAGA AAGCTCATGA 2200
CCAATTTCTA GCCGACCTTG CGAGCATTCT ACCGAGTAAC ACCACACCGC 2250
TCATTGTCAG TGATGCTGGC TTTAAAGTGC CATGGTATAA ATCCGTTGAG 2300
AAGCTGGGTT GGTACTGGTT AAGTCGAGTA AGAGGAAAAG TACAATATGC 2350
AGACCTAGGA GCGGAAAACT GGAAACCTAT CAGCAACTTA CATGATATGT 2400
CATCTAGTCA CTCAAAGACT TTAGGCTATA AGAGGCTGAC TAAAAGCAAT 2450
CCAATCTCAT GCCAAATTCT ATTGTATAAA TCTCGCTCTA AAGGCCGAAA 2500
AAATCAGCGC TCGACACGGA CTCATTGTCA CCACCCGTCA CCTAAAATCT 2550
ACTCAGCGTC GGCAAAGGAG CCATGGGTTC TAGCAACTAA CTTACCTGTT 2600
GAAATTCGAA CACCCAAACA ACTTGTTAAT ATCTATTCGA AGCGAATGCA 2650
GATTGAAGAA ACCTTCCGAG ACTTGAAAAG TCCTGCCTAC GGACTAGGCC 2700
TACGCCATAG CCGAACGAGC AGCTCAGAGC GTTTTGATAT CATGCTGCTA 2750
ATCGCCCTGA TGCTTCAACT AACATGTTGG CTTGCGGGCG TTCATGCTCA 2800
GAAACAAGGT TGGGACAAGC ACTTCCAGGC TAACACAGTC AGAAATCGAA 2850
ACGTACTCTC AACAGTTCGC TTAGGCATGG AAGTTTTGCG GCATTCTGGC 2900
TACACAATAA CAAGGGAAGA CTTACTCGTG GCTGCAACCC TACTAGCTCA 2950
AAATTTATTC ACACATGGTT ACGCTTTGGG GAAATTATGA TAATGATCCA 3000
GATCACTTCT GGCTAATAAA AGATCAGAGC TCTAGAGATC TGTGTGTTGG 3050
TTTTTTGTGG ATCTGCTGTG CCTTCTAGTT GCCAGCCATC TGTTGTTTGC 3100
CCCTCCCCCG TGCCTTCCTT GACCCTGGAA GGTGCCACTC CCACTGTCCT 3150
TTCCTAATAA AATGAGGAAA TTGCATCGCA TTGTCTGAGT AGGTGTCATT 3200
CTATTCTGGG GGGTGGGGTG GGGCAGCACA GCAAGGGGGA GGATTGGGAA 3250
GACAATAGCA GGCATGCTGG GGATGCGGTG GGCTCTATGG GTACCTCTCT 3300
CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCGGTAC CTCTCTCTCT 3350
CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CGGTACCAGG TGCTGAAGAA 3400
TTGACCCGGT GACCAAAGGT GCCTTTTATC ATCACTTTAA AAATAAAAAA 3450
CAATTACTCA GTGCCTGTTA TAAGCAGCAA TTAATTATGA TTGATGCCTA 3500
CATCACAACA AAAACTGATT TAACAAATGG TTGGTCTGCC TTAGAAAGTA 3550
```

TATTTGAACA TTATCTTGAT TATATTATTG ATAATAATAA AAACCTTATC 3600
CCTATCCAAG AAGTGATGCC TATCATTGGT TGGAATGAAC TTGAAAAAAA 3650
TTAGCCTTGA ATACATTACT GGTAAGGTAA ACGCCATTGT CAGCAAATTG 3700
ATCCAAGAGA ACCAACTTAA AGCTTTCCTG ACGGAATGTT AATTCTCGTT 3750
GACCCTGAGC ACTGATGAAT CCCCTAATGA TTTTGGTAAA AATCATTAAG 3800
TTAAGGTGGA TACACATCTT GTCATATGAT CCCGGTAATG TGAGTTAGCT 3850
CACTCATTAG GCACCCCAGG CTTTACACTT TATGCTTCCG GCTCGTATGT 3900
TGTGTGGAAT TGTGAGCGGA TAACAATTTC ACACAGGAAA CAGCTATGAC 3950
CATGATTACG CCAAGCGCGC AATTAACCCT CACTAAAGGG AACAAAAGCT 4000
GGAGCTCCAC CGCGGTGGCG GCCGCTCTAG AACTAGTGGA TCCCCCGGGC 4050
TGCAGGAATT CGATATCAAG CTTATCGATA CCGCTGACCT CGAGGGGGGG 4100
CCCGGTACCC AATTCGCCCT ATAGTGAGTC GTATTACGCG CGCTCACTGG 4150
CCGTCGTTTT ACAACGTCGT GACTGGGAAA ACCCTGGCGT TACCCAACTT 4200
AATCGCCTTG CAGCACATCC CCCTTTCGCC AGCTGGCGTA ATAGCGAAGA 4250
GGCCCGCACC GATCGCCCTT CCCAACAGTT GCGCAGCCTG AATGGCGAAT 4300
GGAAATTGTA AGCGTTAATA TTTTGTTAAA ATTCGCGTTA AATTTTTGTT 4350
AAATCAGCTC ATTTTTTAAC CAATAGGCCG AAATCGGCAA AATCCCTTAT 4400
AAATCAAAAG AATAGACCGA GATAGGGTTG AGTGTTGTTC CAGTTTGGAA 4450
CAAGAGTCCA CTATTAAAGA ACGTGGACTC CAACGTCAAA GGGCGAAAAA 4500
CCGTCTATCA GGGCGATGGC CCACTACTCC GGGATCATAT GACAAGATGT 4550
GTATCCACCT TAACTTAATG ATTTTTACCA AAATCATTAG GGGATTCATC 4600
AGTGCTCAGG GTCAACGAGA ATTAACATTC CGTCAGGAAA GCTTATGATG 4650
ATGATGTGCT TAAAAACTTA CTCAATGGCT GGTTATGCAT ATCGCAATAC 4700
ATGCGAAAAA CCTAAAAGAG CTTGCCGATA AAAAAGGCCA ATTTATTGCT 4750
ATTTACCGCG GCTTTTTATT GAGCTTGAAA GATAAATAAA ATAGATAGGT 4800
TTTATTTGAA GCTAAATCTT CTTTATCGTA AAAAATGCCC TCTTGGGTTA 4850
TCAAGAGGGT CATTATATTT CGCGGAATAA CATCATTTGG TGACGAAATA 4900
ACTAAGCACT TGTCTCCTGT TTACTCCCCT GAGCTTGAGG GGTTAACATG 4950
AAGGTCATCG ATAGCAGGAT AATAATACAG TAAAACGCTA AACCAATAAT 5000
CCAAATCCAG CCATCCCAAA TTGGTAGTGA ATGATTATAA ATAACAGCAA 5050
ACAGTAATGG GCCAATAACA CCGGTTGCAT TGGTAAGGCT CACCAATAAT 5100
CCCTGTAAAG CACCTTGCTG ATGACTCTTT GTTTGGATAG ACATCACTCC 5150
CTGTAATGCA GGTAAAGCGA TCCCACCACC AGCCAATAAA ATTAAAACAG 5200
GGAAAACTAA CCAACCTTCA GATATAAACG CTAAAAAGGC AAATGCACTA 5250
CTATCTGCAA TAAATCCGAG CAGTACTGCC GTTTTTTCGC CCATTTAGTG 5300

GCTATTCTTC CTGCCACAAA GGCTTGGAAT ACTGAGTGTA AAAGACCAAG 5350
ACCCGTAATG AAAAGCCAAC CATCATGCTA TTCATCATCA CGATTTCTGT 5400
AATAGCACCA CACCGTGCTG GATTGGCTAT CAATGCGCTG AAATAATAAT 5450
CAACAAATGG CATCGTTAAA TAAGTGATGT ATACCGATCA GCTTTTGTTC 5500
CCTTTAGTGA GGGTTAATTG CGCGCTTGGC GTAATCATGG TCATAGCTGT 5550
TTCCTGTGTG AAATTGTTAT CCGCTCACAA TTCCACACAA CATACGAGCC 5600
GGAAGCATAA AGTGTAAAGC CTGGGGTGCC TAATGAGTGA GCTAACTCAC 5650
ATTAATTGCG TTGCGCTCAC TGCCCGCTTT CCAGTCGGGA AACCTGTCGT 5700
GCCAGCTGCA TTAATGAATC GGCCAACGCG CGGGGAGAGG CGGTTTGCGT 5750
ATTGGGCGCT CTTCCGCTTC CTCGCTCACT GACTCGCTGC GCTCGGTCGT 5800
TCGGCTGCGG CGAGCGGTAT CAGCTCACTC AAAGGCGGTA ATACGGTTAT 5850
CCACAGAATC AGGGGATAAC GCAGGAAAGA ACATGTGAGC AAAAGGCCAG 5900
CAAAAGGCCA GGAACCGTAA AAGGCCGCGT TGCTGGCGTT TTTCCATAG 5950
GCTCCGCCCC CCTGACGAGC ATCACAAAAA TCGACGCTCA AGTCAGAGGT 6000
GGCGAAACCC GACAGGACTA TAAAGATACC AGGCGTTTCC CCCTGGAAGC 6050
TCCCTCGTGC GCTCTCCTGT TCCGACCCTG CCGCTTACCG GATACCTGTC 6100
CGCCTTTCTC CCTTCGGGAA GCGTGGCGCT TTCTCATAGC TCACGCTGTA 6150
GGTATCTCAG TTCGGTGTAG GTCGTTCGCT CCAAGCTGGG CTGTGTGCAC 6200
GAACCCCCCG TTCAGCCCGA CCGCTGCGCC TTATCCGGTA ACTATCGTCT 6250
TGAGTCCAAC CCGGTAAGAC ACGACTTATC GCCACTGGCA GCAGCCACTG 6300
GTAACAGGAT TAGCAGAGCG AGGTATGTAG GCGGTGCTAC AGAGTTCTTG 6350
AAGTGGTGGC CTAACTACGG CTACACTAGA AGGACAGTAT TTGGTATCTG 6400
CGCTCTGCTG AAGCCAGTTA CCTTCGGAAA AAGAGTTGGT AGCTCTTGAT 6450
CCGGCAAACA AACCACCGCT GGTAGCGGTG GTTTTTTTGT TTGCAAGCAG 6500
CAGATTACGC GCAGAAAAAA AGGATCTCAA GAAGATCCTT TGATCTTTTC 6550
TACGGGGTCT GACGCTCAGT GGAACGAAAA CTCACGTTAA GGGATTTTGG 6600
TCATGAGATT ATCAAAAAGG ATCTTCACCT AGATCCTTTT AAATTAAAAA 6650
TGAAGTTTTA AATCAATCTA AAGTATATAT GAGTAAACTT GGTCTGACAG 6700
TTACCAATGC TTAATCAGTG AGGCACCTAT CTCAGCGATC TGTCTATTTC 6750
GTTCATCCAT AGTTGCCTGA CTCCCCGTCG TGTAGATAAC TACGATACGG 6800
GAGGGCTTAC CATCTGGCCC CAGTGCTGCA ATGATACCGC GAGACCCACG 6850
CTCACCGGCT CCAGATTTAT CAGCAATAAA CCAGCCAGCC GGAAGGGCCG 6900
AGCGCAGAAG TGGTCCTGCA ACTTTATCCG CCTCCATCCA GTCTATTAAT 6950
TGTTGCCGGG AAGCTAGAGT AAGTAGTTCG CCAGTTAATA GTTTGCGCAA 7000
CGTTGTTGCC ATTGCTACAG GCATCGTGGT GTCACGCTCG TCGTTTGGTA 7050

TGGCTTCATT CAGCTCCGGT TCCCAACGAT CAAGGCGAGT TACATGATCC 7100
CCCATGTTGT GCAAAAAAGC GGTTAGCTCC TTCGGTCCTC CGATCGTTGT 7150
CAGAAGTAAG TTGGCCGCAG TGTTATCACT CATGGTTATG GCAGCACTGC 7200
ATAATTCTCT TACTGTCATG CCATCCGTAA GATGCTTTTC TGTGACTGGT 7250
GAGTACTCAA CCAAGTCATT CTGAGAATAG TGTATGCGGC GACCGAGTTG 7300
CTCTTGCCCG GCGTCAATAC GGGATAATAC CGCGCCACAT AGCAGAACTT 7350
TAAAAGTGCT CATCATTGGA AAACGTTCTT CGGGGCGAAA ACTCTCAAGG 7400
ATCTTACCGC TGTTGAGATC CAGTTCGATG TAACCCACTC GTGCACCCAA 7450
CTGATCTTCA GCATCTTTTA CTTTCACCAG CGTTTCTGGG TGAGCAAAAA 7500
CAGGAAGGCA AAATGCCGCA AAAAAGGGAA TAAGGGCGAC ACGGAAATGT 7550
TGAATACTCA TACTCTTCCT TTTTCAATAT TATTGAAGCA TTTATCAGGG 7600
TTATTGTCTC ATGAGCGGAT ACATATTTGA ATGTATTTAG AAAAATAAAC 7650
AAATAGGGGT TCCGCGCACA TTTCCCCGAA AAGTGCCAC 7689

SEQ ID NO:58(a Kozak sequence)
ACCATGG

SEQ ID NO:59 (a Kozak sequence)
ACCATGT

SEQ ID NO:60 (a Kozak sequence)
AAGATGT

SEQ ID NO:61 (a Kozak sequence)
ACGATGA

SEQ ID NO:62 (a Kozak sequence)
AAGATGG

SEQ ID NO:63 (a Kozak sequence)
GACATGA

SEQ ID NO:64 (a Kozak sequence)
ACCATGA

SEQ ID NO:65 (a Kozak sequence)
ACCATGT

SEQ ID NO:66 (conalbumin polyA)
tctgccattg ctgcttcctc tgcccttcct cgtcactctg aatgtggctt cttcgctact
gccacagcaa gaaataaaat ctcaacatct aaatgggttt cctgaggttt ttcaagagtc
gttaagcaca ttccttcccc agcacccctt gctgcaggcc agtgccaggc accaacttgg
ctactgctgc ccatgagaga aatccagttc aatattttcc aaagcaaaat ggattacata
tgccctagat cctgattaac aggcgtttgt attatctagt gctttcgctt cacccagatt
atcccattgc ctccc

SEQ ID NO:67 (synthetic polyA)
GGCGCCTGGATCCAGATCACTTCTGGCTAATAAAAGATCAGAGCTCTAGAGATCTGTGTGTTGGTTTTTTG
TGGATCTGCTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGG
AAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCAT
TCTATTCTGGGGGGTGGGGTGGGGCAGCACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGG
GGATGCGGTGGGCTCTATGGGTACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCGGTACC
TCTCTC

SEQ ID NO:68 (avian optimized polyA)
ggggatcgc tctagagcga tccgggatct cgggaaaagc gttggtgacc aaaggtgcct
tttatcatca ctttaaaaat aaaaaacaat tactcagtgc ctgttataag cagcaattaa
ttatgattga tgcctacatc acaacaaaaa ctgatttaac aaatggttgg tctgccttag
aaagtatatt tgaacattat cttgattata ttattgataa taataaaaac cttatcccta
tccaagaagt gatgcctatc attggttgga atgaacttga aaaaaattag ccttgaatac
attactggta aggtaaacgc cattgtcagc aaattgatcc aagagaacca a

SEQ ID NO:69 (vitellogenin promoter)
TGAATGTGTT CTTGTGTTAT CAATATAAAT CACAGTTAGT GATGAAGTTG GCTGCAAGCC
TGCATCAGTT CAGCTACTTG GCTGCATTTT GTATTTGGTT CTGTAGGAAA TGCAAAAGGT
TCTAGGCTGA CCTGCACTTC TATCCCTCTT GCCTTACTGC TGAGAATCTC TGCAGGTTTT
AATTGTTCAC ATTTTGCTCC CATTTACTTT GGAAGATAAA ATATTTACAG AATGCTTATG

AAACCTTTGT TCATTTAAAA ATATTCCTGG TCAGCGTGAC CGGAGCTGAA AGAACACATT
GATCCCGTGA TTTCAATAAA TACATATGTT CCATATATTG TTTCTCAGTA GCCTCTTAAA
TCATGTGCGT TGGTGCACAT ATGAATACAT GAATAGCAAA GGTTTATCTG GATTACGCTC
TGGCCTGCAG GAATGGCCAT AAACCAAAGC TGAGGGAAGA GGGAGAGTAT AGTCAATGTA
GATTATACTG ATTGCTGATT GGGTTATTAT CAGCTAGATA ACAACTTGGG TCAGGTGCCA
GGTCAACATA ACCTGGGCAA AACCAGTCTC ATCTGTGGCA GGACCATGTA CCAGCAGCCA
GCCGTGACCC AATCTAGGAA AGCAAGTAGC ACATCAATTT TAAATTTATT GTAAATGCCG
TAGTAGAAGT GTTTTACTGT GATACATTGA AACTTCTGGT CAATCAGAAA AAGGTTTTTT
ATCAGAGATG CCAAGGTATT ATTTGATTTT CTTTATTCGC CGTGAAGAGA ATTTATGATT
GCAAAAAGAG GAGTGTTTAC ATAAACTGAT AAAAAACTTG AGGAATTCAG CAGAAAACAG
CCACGTGTTC CTGAACATTC TTCCATAAAA GTCTCACCAT GCCTGGCAGA GCCCTATTCA
CCTTCGCT

SEQ ID NO:70 (fragment of ovalbumin promoter - chicken)
GAGGTCAGAAT GGTTTCTTTA CTGTTTGTCA ATTCTATTAT TTCAATACAG
AACAATAGCT TCTATAACTG AAATATATTT GCTATTGTAT ATTATGATTG
TCCCTCGAAC CATGAACACT CCTCCAGCTG AATTTCACAA TTCCTCTGTC
ATCTGCCAGG CCATTAAGTT ATTCATGGAA GATCTTTGAG GAACACTGCA
AGTTCATATC ATAAACACAT TTGAAATTGA GTATTGTTTT GCATTGTATG
GAGCTATGTT TTGCTGTATC CTCAGAAAAA AAGTTTGTTA TAAAGCATTC
ACACCCATAA AAAGATAGAT TTAAATATTC CAGCTATAGG AAAGAAAGTG
CGTCTGCTCT TCACTCTAGT CTCAGTTGGC TCCTTCACAT GCATGCTTCT
TTATTTCTCC TATTTTGTCA AGAAAATAAT AGGTCACGTC TTGTTCTCAC
TTATGTCCTG CCTAGCATGG CTCAGATGCA CGTTGTAGAT ACAAGAAGGA
TCAAATGAAA CAGACTTCTG GTCTGTTACT ACAACCATAG TAATAAGCAC
ACTAACTAAT AATTGCTAAT TATGTTTTCC ATCTCTAAGG TTCCCACATT
TTTCTGTTTT CTTAAAGATC CCATTATCTG GTTGTAACTG AAGCTCAATG
GAACATGAGC AATATTTCCC AGTCTTCTCT CCCATCCAAC AGTCCTGATG
GATTAGCAGA ACAGGCAGAA AACACATTGT TACCCAGAAT TAAAAACTAA
TATTTGCTCT CCATTCAATC CAAAATGGAC CTATTGAAAC TAAAATCTAA
CCCAATCCCA TTAAATGATT TCTATGGCGT CAAAGGTCAA ACTTCTGAAG
GGAACCTGTG GGTGGGTCAC AATTCAGGCT ATATATTCCC CAGGGCTCAG

SEQ ID NO:71 (chicken ovalbumin ehancer)
ccgggctgca gaaaaatgcc aggtggacta tgaactcaca tccaaaggag

cttgacctga tacctgattt tcttcaaact ggggaaacaa cacaatccca caaaacagct
cagagagaaa ccatcactga tggctacagc accaaggtat gcaatggcaa tccattcgac
attcatctgt gacctgagca aaatgattta tctctccatg aatggttgct tctttccctc
atgaaaaggc aatttccaca ctcacaatat gcaacaaaga caaacagaga acaattaatg
tgctccttcc taatgtcaaa attgtagtgg caaagaggag aacaaaatct caagttctga
gtaggtttta gtgattggat aagaggcttt gacctgtgag ctcacctgga cttcatatcc
ttttggataa aaagtgcttt tataactttc aggtctccga gtctttattc atgagactgt
tggtttaggg acagacccac aatgaaatgc ctggcatagg aaagggcagc agagccttag
ctgacctttt cttgggacaa gcattgtcaa acaatgtgtg acaaaactat ttgtactgct
ttgcacagct gtgctgggca gggcaatcca ttgccaccta tcccaggtaa ccttccaact
gcaagaagat tgttgcttac tctctctaga

SEQ ID NO:72 (5' untranslated region)
GTGGATCAACATACAGCTAGAAAGCTGTATTGCCTTTAGCACTCAAGCTCAAAAGACAACTCAGAGTTCAC
C

SEQ ID NO:73 (putative cap site)
ACATACAGCTAG AAAGCTGTAT TGCCTTTAGC ACTCAAGCTC AAAAGACAAC TCAGAGTTCA

SEQ ID NO:74 (Chicken Ovalbumin Signal Sequence)
ATG GGCTCCATCG GCGCAGCAAG CATGGAATTT TGTTTTGATG TATTCAAGGA GCTCAAAGTC
CACCATGCCA ATGAGAACAT CTTCTACTGC CCCATTGCCA TCATGTCAGC TCTAGCCATG
GTATACCTGG GTGCAAAAGA CAGCACCAGG ACACAGATAA ATAAGGTTGT TCGCTTTGAT
AAACTTCCAG GATTCGGAGA CAGTATTGAA GCTCAGTGTG GCACATCTGT AAACGTTCAC
TCTTCACTTA GAGACATCCT CAACCAAATC ACCAAACCAA ATGATGTTTA TTCGTTCAGC
CTTGCCAGTA GACTTTATGC TGAAGAGAGA TACCCAATCC TGCCAGAATA CTTGCAGTGT
GTGAAGGAAC TGTATAGAGG AGGCTTGGAA CCTATCAACT TTCAAACAGC TGCAGATCAA
GCCAGAGAGC TCATCAATTC CTGGGTAGAA AGTCAGACAA ATGGAATTAT CAGAAATGTC
CTTCAGCCAA GCTCCGTGGA TTCTCAAACT GCAATGGTTC TGGTTAATGC CATTGTCTTC
AAAGGACTGT GGGAGAAAAC ATTTAAGGAT GAAGACACAC AAGCAATGCC TTTCAGAGTG
ACTGAGCAAG AAAGCAAACC TGTGCAGATG ATGTACCAGA TTGGTTTATT TAGAGTGGCA
TCAATGGCTT CTGAGAAAAT GAAGATCCTG GAGCTTCCAT TTGCCAGTGG GACAATGAGC
ATGTTGGTGC TGTTGCCTGA TGAAGTCTCA GGCCTTGAGC AGCTTGAGAG TATAATCAAC
TTTGAAAAAC TGACTGAATG GACCAGTTCT AATGTTATGG AAGAGAGGAA GATCAAAGTG
TACTTACCTC GCATGAAGAT GGAGGAAAAA TACAACCTCA CATCTGTCTT AATGGCTATG

GGCATTACTG ACGTGTTTAG CTCTTCAGCC AATCTGTCTG GCATCTCCTC AGCAGAGAGC
CTGAAGATAT CTCAAGCTGT CCATGCAGCA CATGCAGAAA TCAATGAAGC AGGCAGAGAG
GTGGTAGGGT CAGCAGAGGC TGGAGTGGAT GCTGCAAGCG TCTCTGAAGA ATTTAGGGCT
GACCATCCAT TCCTCTTCTG TATCAAGCAC ATCGCAACCA ACGCCGTTCT CTTCTTTGGC
AGATGTGTTT CCCCT

SEQ ID NO:75 (Chicken Ovalbumin Signal Sequence-shortened approx. 50bp)
ATG GGCTCCATCG GCGCAGCAAG CATGGAATTT TGTTTTGATG TATTCAAG

SEQ ID NO:76 (Chicken Ovalbumin Signal Sequence-shortened approx. 100bp)
ATG GGCTCCATCG GCGCAGCAAG CATGGAATTT TGTTTTGATG TATTCAAGGA GCTCAAAGTC
CACCATGCCA ATGAGAACAT CTTCTACTGC CCCATTGCCA TC

SEQ ID NO:77 (vitellogenin targeting sequence)
ATGAGGGGGATCATACTGGCATTAGTGCTCACCCTTGTAGGCAGCCAGAAGTTTGACATTGGT

SEQ ID NO:78 (p146 protein)
KYKKALKKLAKLL

SEQ ID NO:79 (p146 coding sequence)
AAATACAAAAAAGCACTGAAAAAACTGGCAAAACTGCTG

SEQ ID NO:80 (pro-insulin sequence)
TTTGTGAACCAACACCTGTGCGGCTCACACCTGGTGGAAGCTCTCTACCTAGTGTGCGGGGAACGAGGCTT
CTTCTACACACCCAAGACCCGCCGGGAGGCAGAGGACCTGCAGGTGGGGCAGGTGGAGCTGGGCGGGGGCC
CTGGTGCAGGCAGCCTGCAGCCCTTGGCCCTGGAGGGGTCCCTGCAGAAGCGTGGCATTGTGGAACAATGC
TGTACCAGCATCTGCTCCCTCTACCAGCTGGAGAACTCTGCAACTAG

SEQ ID NO:81 (spacer)
$(GPGG)_x$

SEQ ID NO:82 (spacer)
GPGGGPGGGPGG

SEQ ID NO:83 (spacer)
GGGGSGGGGSGGGGS

SEQ ID NO:84 (spacer)
GGGGSGGGGSGGGGSGGGGS

SEQ ID NO:85 (repeat domain in TAG spacer sequence)
Pro Ala Asp Asp Ala

SEQ ID NO:86 (TAG spacer sequence)
Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp

SEQ ID NO:87 (gp41 epitope)
Ala Thr Thr Cys Ile Leu Lys Gly Ser Cys Gly Trp Ile Gly Leu Leu

SEQ ID NO:88 (polynucleotide sequence encoding gp41 epitope)
Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Thr Thr Cys Ile Leu Lys Gly Ser Cys Gly Trp Ile Gly Leu Leu Asp Asp Asp Asp Lys

SEQ ID NO:89 (enterokinase cleavage site)
DDDDK

SEQ ID NO:90 (TAG sequence)
Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Thr Thr Cys Ile Leu Lys Gly Ser Cys Gly Trp Ile Gly Leu Leu Asp Asp Asp Asp Lys

SEQ ID NO:91 (altered transposase Hef forward primer)
ATCTCGAGACCATGTGTGAACTTGATATTTTACATGATTCTCTTTACC

SEQ ID NO:92 (altered transposase Her reverse primer)
GATTGATCATTATCATAATTTCCCCAAAGCGTAACC

SEQ ID NO:93 (Xho I restriction site)
CTCGAG

SEQ ID NO:94 (Bcl I restriction site)
TGATCA

SEQ ID NO:95 (CMVf-NgoM IV primer)
TTGCCGGCATCAGATTGGCTAT

SEQ ID NO:96 (Syn-polyAr-BstE II primer)
AGAGGTCACCGGGTCAATTCTTCAGCACCTGGTA

SEQ ID NO:97 (pTnMOD (CMV-prepro-HCPro-Lys-CPA)

```
   1 ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga
  61 ccgctacact tgccagcgcc ctagcgcccg ctcctttcgc tttcttccct tcctttctcg
 121 ccacgttcgc cggcatcaga ttggctattg gccattgcat acgttgtatc catatcataa
 181 tatgtacatt tatattggct catgtccaac attaccgcca tgttgacatt gattattgac
 241 tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg
 301 cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt
 361 gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca
 421 atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc
 481 aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta
 541 catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac
 601 catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg
 661 atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg
 721 ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt
 781 acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg
 841 ccatccacgc tgttttgacc tccatagaag acaccgggac cgatccagcc tccgcggccg
 901 ggaacggtgc attggaacgc ggattccccg tgccaagagt gacgtaagta ccgcctatag
 961 actctatagg cacacccctt tggctcttat gcatgctata ctgtttttgg cttggggcct
1021 atacacccgc gcttccttat gctataggtg atggtatagc ttagcctata ggtgtgggtt
1081 attgaccatt attgaccact cccctattgg tgacgatact ttccattact aatccataac
1141 atggctcttt gccacaacta tctctattgg ctatatgcca atactctgtc cttcagagac
```

```
1201 tgacacggac tctgtatttt tacaggatgg ggtcccattt attatttaca aattcacata
1261 tacaacaacg ccgtcccccg tgcccgcagt ttttattaaa catagcgtgg gatcttccacg
1321 cgaatctcgg gtacgtgttc cggacatggg ctcttctccg gtagcggcgg agcttccaca
1381 tccgagccct ggtcccatgc tcatggtcgc tcggcagctc cttgctccta
1441 acagtggagg ccagacttag gcacagcaca atgcccacca ccaccagtgt gccgcacaag
1501 gccgtggcgg taggtatgt gtctgaaaat gagcgtggag attgggctcg cacggctgac
1561 gcagatggaa gacttaaggc agcggcagaa gaagatgcag gcagctgagt tgttgtattc
1621 tgataagagt cagaggtaac tcccgttgcg gtgctgttaa cggtggaggg cagtgtagtc
1681 tgagcagtac tcgttgctgc cgcgcgcgcc accagacata atagctgaca gactaacaga
1741 ctgttccttt ccatgggtct ttcctgcagt caccgtcgga ccatgtgtga acttgatatt
1801 ttacatgatt ctccttacca atcctgcccc gaattacact taaaacgact caacagctta
1861 acgttggctt gccacgcatt acttgactgt aaaactctca ctcttaccga acttggccgt
1921 aacctgccaa ccaaagcgag aacaaaacat aacatcaaac gaatcgaccg attgttaggt
1981 aatcgtcacc tccacaaaga gcgactcgct gtatacgtgt ggcatgctag ctttatctgt
2041 tcgggcaata cgatgcccat tgtacttgtt gactggtctg atattcgtga gcaaaaacga
2101 cttatggtat tgcgagcttc agtcgcacta cacggtcgtt ctgttactct ttatgagaaa
2161 gcgttcccgc tttcagagca atgttcaaag aaagctcatg accaatttct agccgacctt
2221 gcgagcattc taccgagtaa caccacaccg ctcattgtca gtgatgctgg ctttaaagtg
2281 ccatggtata aatccgttga gaagctgggt tggtactggt taagtcgagt aagaggaaaa
2341 gtacaatatg cagacctagg agcggaaaac tggaaaccta tcagcaactt acatgatatg
2401 tcatctagtc actcaaagac tttaggctat aagaggctga ctaaaagcaa tccaatctca
2461 tgccaaattc tattgtataa atctcgctct aaaggccgaa aaaatcagcg ctcgacacgg
2521 actcattgtc accacccgtc acctaaaatc tactcagcgt cggcaaagga gccatgggtt
2581 ctagcaacta acttacctgt tgaaattcga acacccaaac aacttgttaa tatctattcg
2641 aagcgaatgc agattgaaga aaccttccga gacttgaaaa gtcctgccta cggaccaggc
2701 ctacgccata gccgaacgag cagctcagag cgttttgata tcatgctgct aatcgccctg
2761 atgcttcaac taacatgttg gcttgcgggc gttcatgctc agaaacaagg ttgggacaag
2821 cacttccagg ctaacacagt cagaaatcga aacgtactct caacagttcg cttaggcatg
2881 gaagttttgc ggcattctgg ctacacaata acaagggaag acttactcgt ggctgcaacc
2941 ctactagctc aaaatttatt cacacatggt tacgctttgg ggaaattatg ataatgatcc
3001 agatcacttc tggctaataa aagatcagag ctctagagat ctgtgtgttg gttttttgtg
3061 gatctgctgt gccttctagt tgccagccat ctgtttgttg cccctccccc gtgccttcct
3121 tgaccctgga aggtgccact cccactgtcc ttcctaata aaatgaggaa attgcatcgc
3181 attgtcctgag taggtgtcat tctattctgg ggggtggggt ggggcagcac agcaaggggg
3241 aggattggga agacaatagc aggcatgctg gggatgcggt gggcttatg ggtaccttc
3301 tctctctctc tctctctctc tctctctctc cctctctctc tctctctctc
3361 tctctctctc tctctctctc tcggtaccag gtgctgaaga attgacccgg tgaccaaagg
3421 tgcctttat catcacttta aaaataaaaa acaattactc agtgccttgtt ataagcagca
3481 attaattatg attgatgcct acatcacaac aaaaactgat ttaacaaatg gttggtctgc
```

3541 cttagaaagt atatttgaac attatcttga ttatattatt gataataata aaaaccttat
3601 ccctatccaa gaagtgatgc ctatcattgg ttggaatgaa cttgaaaaaa attagccttg
3661 aatacattac tggtaaggta aacgccattg tcagcaaatt gatccaagag aaccaactta
3721 aagctttcct gacggaatgt taattctcgt tgaccctgag cactgatgaa tcccctaatg
3781 atttggtaa aaatcattaa gttaaggtgg atacacatct tgtcatatga tcccggtaat
3841 gtgagttagc tcactcatta ggcacccaag gctttacact ttatgcttcc ggctcgtatg
3901 ttgtgtggaa ttgtgagcgg ataaacaatt cacacaggaa acagctatga ccatgattac
3961 gccaagcgcg caattaaccc tcactaaagg gaacaaaagc tggagctcca ccgcggtggc
4021 ggccgctcta gaactagtgg atcccccggg ctgcaggaat tcgatatcaa gcttatcgat
4081 accgctgacc tcgagcatca gattggctat ataccttgta tccatatcat
4141 aatatgtaca tttatattgg ctcatgtcca acattaccgc catgttgaca ttgattattg
4201 actagttatt aatagtaatc aattacgggg tcattagttc atagcccata tatggagttc
4261 cgcgttacat aacttacggt aaatggcccg cgctggctgac cgcccaacga ccccgccca
4321 ttgacgtcaa taatgacgta tgttcccata gtaacgccaa tagggacttt ccattgacgt
4381 caatgggtgg agtatttacg gtaaactgcc cacttggcag tacatcaagt gtatcatatg
4441 tcaagtacgc cccctattga cgtcaatgac ggtaaatggc ccgcctggca ttatgcccag
4501 tacatgacct tatgggactt tcctacttgg cagtacatct acgtattagt catcgctatt
4561 accatggtga tgcggttttg gcagtacatc aatgggcgtg gatagcggtt tgactcacgg
4621 ggatttccaa gtctcaccc cattgacgtc aatgggagtt tgttttggca ccaaaatcaa
4681 cgggactttc caaaatgtcg taacaactcc gccccattga cgcaaatggg cggtaggcgt
4741 gtacggtggg aggtctatat aagcagagct cgtttagtga accgtcagat cgcctggaga
4801 cgccatccac gctgttttga cctccataga agacaccggg accgatccag cctccgcggc
4861 cgggaacggt gcattggaac gcggattccc cgtgccaaga gtgacgtaag taccgcctat
4921 agactctata ggcacacccc tttggctctt atgcatgcta tactgttttt ggcttsggc
4981 ctatacaccc ccgcttcctt atgctatagg tgatggtata gcttagccta taggtgtggg
5041 ttattgacca ttattgacca ctcccctatt ggtgacgata ctttccatta ctaatccata
5101 acatggctct ttgccacaac tatcttcatt ggctatatgc caatacttg tccttcagag
5161 actgacacgg actctgtatt tttacaggat ggggtcccat ttattattta caaattcaca
5221 tatacaacaa cgccgtcccc cgtgcccgca gttttatta aacatagcgt gggatcttcca
5281 cgcgaatctc gggtacgtgt tccggacatg ggctcttctc cggtagcggc ggagcttcca
5341 catccgagcc ctggtcccat gcctccagcg gctccatggtc gctcggcagc tccttgctcc
5401 taacagtgga ggccagactt aggcacagca caatgcccac caccaccagt gtgccgcaca
5461 aggccgtggc ggtagggtat gtgtctgaaa atgagcgtgg agattgggct cgcacggctg
5521 acgcagatgg aagacttaag gcagcggcag aagaagatgc aggcagctga gttgttgtat
5581 tctgataaga gtcagaggta actcccgttg cggtgctgtt aacggtggag ggcagtgtag
5641 tctgagcagt actcgttgct gccgcgcgcg ccaccagaca taatagctga cagactaaca
5701 gactgttcct tcccatgggt ctttctgca gtccaccgtcg gatcaatcat tcatctcgtg
5761 acttcttcgt gtgtggtgtt tacctatata tctaaattta atatttcgtt tattaaaatt
5821 taatatattt cgacgatgaa tttctcaagg atatttttct tcgtgttcgc tttggttctg

5881 gctttgtcaa cagtttcggc tgcgccagag ccgaaaggta cccaggtgca gctgcaggag
5941 tcgggggagg gcttggtaaa gccgggggcg tcccttagag tctcctgtgc agcctctgga
6001 ttcactttca gaaacgcctg gatgagctgg gtccgccagg ctccagggaa gggctggag
6061 tgggtcggcc gtattaaaag caaaattgat ggtgggacaa cagactatgc tgcacccgtg
6121 aaaggcagat tcaccatctc aagagatgat tcaaaaaaca cgttatatct gcaaatgaat
6181 agcctgaaag ccgaggacac agccgtatat tactgtacca cgggattat gataacattt
6241 gggggagtta tccctccccc gaattgggcc cagggaaccc tggtcaccgt ctcctcagcc
6301 tccaccaagg gcccatcggt cttcccccctg gcaccctcct ccaagagcac ctctgggggc
6361 acagcgggccc tgggctgcct ggtcaaggac tacttccccg aaccggtgac ggtgtcgtgg
6421 aactcaggcg ccctgaccag cggcgtgcac acctttccgg ctgtcctaca gtccttcagga
6481 ctctacttcc ttagcaacgt ggtgaccgtg ccctccagca gcttgggcac ccagacctac
6541 atctgcaacg tgaatcacaa gcccagcaac accaaggtgg acaagaaagt tgagcccaaa
6601 tcttgtgaca aaactcacac atgcccaccg tgcccagcac ctgaactcct gggggggaccg
6661 tcagtcttcc tcttcccccc aaaacccaag gacacccтca tgatctcccg gacccctgag
6721 gtcacatgcg tggtggtgga cgtgagccac gaagaccctg aggtcaagtt caactggtac
6781 gtggacggcg tggaggtgca taatgccaag acaaagccgc gggaggagca gtacaacagc
6841 acgtaccgtg tggtcagcgt cctcaccgtc ctgcaccagg actggctgaa tggcaaggag
6901 tacaagtgca aggtctccaa caaagccctc ccagccccca tcgagaaaac catctccaaa
6961 gccaaagggc agccccgaga accacaggtg tacaccctgc cccatcccg ggatgagctg
7021 accaagaacc aggtcagcct gacctgcctg gtcaaaggct tctatcccag cgacatcgcc
7081 gtggagtggg agagcaatgg gcagccggag aacaactaca agaccacgcc tcccgtgctg
7141 gactccgacg gctccttctt cctctacagc aagctcaccg tggacaagag caggtggcag
7201 caggggaacg tcttctcatg ctccgtgatg catgaggctc tgcacaacca ctacacgcag
7261 aagagcctct ccctgtctcc gggtaaagcg ccagagccga aaaagctttc ctatgaqctg
7321 acacagccac cctcggtgtc agtgtcccca ggacaaacgg ccaggatcac ctgct:tgga
7381 gatgcattgc cagaaaaata tgtttattgg taccagcaga agtcaggcca ggcccctgtg
7441 gtggtcatct atgaggacag caaacgaccc tccgggatcc ctgagagatt ctctggctcc
7501 agctcaggga caatggccac cttgactatc agtggggccc aggtggaaga tgaaggtgac
7561 tactactgtt actcaactga cagcagtggt tatcatagggg aggttgttcag cggagggacc
7621 aagctgaccg tcctaggtca gcccaaggct gcccccctcgg tcactctgtt cccaccctcc
7681 tctgaggagc ttcaagccaa caaggccaca ctggtgtgtc tcataagtga ctcctacccg
7741 ggagccgtga cagtggcctg gaaggcagat agcagccccg tcaaggcggg agtggagacc
7801 accacaccct ccaaacaaag caacaacaag tacgcggcca gcagctacct gagcctgacg
7861 cttgagcagt ggaagtccca caaaagctac agctgccagg tcacgcatga agggagcacc
7921 gtggagaaga cagtggcccc tgcagaatgt tcaccgcgga gggagggaag ggccctttt
7981 gaaggggggag gaaacttcgc gccatgactc ctctcgtgcc cccgcacgg aacactgatg
8041 tgcagagggc cctctgccat tgctgcttcc tctgccctc ctcgtcactc tgaatgtggc
8101 ttctttgcta ctgccacagc aagaaataaa atcttcaacat ctaaatgggt ttcctgagat
8161 tttcaagag tcgttaagca catcccttcc ccagcacccc ttgctgcagg ccagtgccag

8221 gcaccaactt ggctactgct gcccatgaga gaaatccagt tcaatatttt ccaaagcaaa
8281 atggattaca tatgccctag atcctgatta acaggtgttt tgtattatct gtgctttcgc
8341 ttcacccaca ttatcccatt gcctcccctc gactcgaggg ggggcccggt acccaattcg
8401 ccctatagtg agtcgtatta cggcgctcca ctggccgtcg tttacaacg tcgtgactgg
8461 gaaaaccctg gcgttaccca acttaatcgc cttgcagcac atccccctt cgccagctgg
8521 cgtaatagcg aagaggcccg caccgatcgc ccttcccaac agttgcgcag cctgaatggc
8581 gaatggaaat tgtaagcgtt aatattttgt taaaattcgc gttaaatttt tgttaaatca
8641 gctcatttt taaccaatag gccgaaatcg gcaaaatccc ttataaatca aaagaataga
8701 ccgagatagg gttgagtgtt gttccagttt ggaacaagag tccactatta aagaacgtgg
8761 actccaacgt caaagggcga aaaaccgtct atcagggcga tggcccacta cttcgggatc
8821 atatgacaag atgtgtatcc acctaacttt aatgatttct accaaaatca ttagggatt
8881 catcagtgct caggtcaac gagaattaac attccgtcag gaaagcttat gatgatgatg
8941 tgcttaaaaa cttactcaat ggctggttat gcatatcgca atacatgcga aaaacctaaa
9001 agagcttgcc gataaaaaag gccaatttat tgctatttac cgcggctttt tattgagctt
9061 gaaagataaa taaaatagat aggttttatt tgaagctaaa tcttcttat cgtaaaaaat
9121 gcccttttgg gttatcaaga gggtcattat atttcgcgga ataacatcat ttggtgacga
9181 aataactaag cacttgtctc ctgtttactc ccctgagctt gagggttaa catgaaggtc
9241 atcgatagca ggataataat acagtaaaac gctaaaccaa taatccaaat ccagccatcc
9301 caaattggta gtgaatgatt ataaataaca gcaaacagta atgggccaat aacaccggtt
9361 gcattggtaa ggctcaccaa taatccctgt aaagcacctt gctgatgact ctttgtttgg
9421 atagacatca ctccctgtaa tgcaggtaaa gcgatcccac caccagccaa taaaattaaa
9481 acagggaaaa ctaaccaacc ttcagatata aacgctaaaa aggcaaatgc actactatct
9541 gcaataaatc cgagcagtac tgccgtcttt tcgcccattt agtggctatt cttcctgcca
9601 caaaggcttg gaatactgag tgtaaaagac caagacccgt aatgaaaagc caaccatcat
9661 gctattcatc atcacgattt ctgtaatagc accacaccgt gctggattgg ctatcaatgc
9721 gctgaaataa taatcaacaa atggcatcgt taaataagtg atgtataccg atcagcttt
9781 gttcccttta gtgagggtta attggcgcgt tggcgtaatc atggtcatag ctgttcctg
9841 tgtgaaattg ttatccgctc acaattccac acaacatacg agccggaagc ataaagtgta
9901 aagcctgggg tgcctaatga gtgagctaac tcacattaat tgcgttgcgc tcactgcccg
9961 ctttccagtc gggaaacctg tcgtgccagc tgcattaatg aatcggccaa cgcgcgggga
10021 gaggcggttt gcgtattggg cgctcttccg cttcctcgct cactgactcg ctgcgctcgg
10081 tcgttcggct gcggcgagcg gtatcagctc actcaaaggc ggtaatacgg ttatccacag
10141 aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc
10201 gtaaaaaggc cgcgttgctg gcgtttttcc ataggctccg ccccctgac gagcatcaca
10261 aaaatcgacg ctcaagtcag aggtgggcgaa acccgacagg actataaaga taccaggcgt
10321 ttcccctggg aagctccctc gtgcgctctc ctgttccgac cctgccgctt accggatacc
10381 tgtccgcctt tctcccttcg ggaagcgtgg cgctttctca tagctcacgc tgtaggtatc
10441 tcagttcggt gtaggtcgtt cgctccaagc tgggctgtgt gcacgaaccc cccgttcagc
10501 ccgaccgctg cgccttatcc ggtaactatc gtcttgagtc caacccggta agacacgact

```
10561 tatcgccact ggcagcagcc actggtaaca ggattagcag agcgaggtat gtaggcggtg
10621 ctacagagtt cttgaagtgg tggcctaact acggctacac tagaaggaca gtatttggta
10681 tctgcgctct gctgaagcca gttaccttcg gaaaaagagt tggtagctct tgatccggca
10741 aacaaaccac cgctggtagc ggtggttttt ttgtttgcaa gcagcagatt acgcgcagaa
10801 aaaaaggatc tcaagaagat cctttgatct ttttctacggg gtctgacgct cagtggaacg
10861 aaaactcacg ttaagggatt ttggtcatga gattatcaaa aaggatcttc acctagatcc
10921 ttttaaatta aaaatgaagt tttaaatcaa tctaaagtat atatgagtaa acttggtctg
10981 acagttacca atgcttaatc agtgaggcac ctatctcagc gatctgtcta tttcgttcat
11041 ccatagttgc ctgactcccc gtcgtgtaga taactacgat acgggagggc ttaccatctg
11101 gccccagtgc tgcaatgata ccgcgagacc cacgctcacc ggctccagat ttatcagcaa
11161 taaaccagcc agccggaagg gccgagcgca gaagtggtcc tgcaacttta tccgcctcca
11221 tccagtctat taattgttgc cgggaagcta gagtaagtag ttcgccagtt aatagtttgc
11281 gcaacgttgt tgccattgct acaggcatcg tggtgtcacg ctcgtcgttt ggtatggctt
11341 cattcagctc cggttcccaa cgatcaaggc gagttacatg atcccccatg ttgtgcaaaa
11401 aagcggttag ctccttcggt cctccgatcg ttgtcagaag taagttggcc gcagtgttat
11461 cactcatggt tatggcagca ctgcataatt ctcttactgt catgccatcc gtaagatgct
11521 tttctgtgac tggtgagtac tcaaccaagt cattctgaga atagtgtatg cggcgaccga
11581 gttgctcttg cccggcgtca atacgggata ataccgcgcc acatagcaga actttaaaag
11641 tgctcatcat tggaaaacgt tcttcggggc gaaaactctc aaggatctta ccgctgttga
11701 gatccagttc gatgtaaccc actcgtgcac ccaactgatc ttcagcatct tttactttca
11761 ccagcgtttc tgggtgagca aaaacaggaa ggcaaaatgc cgcaaaaaag ggaataaggg
11821 cgacacggaa atgttgaata ctcatactct tcctttttca atattattga agcatttatc
11881 agggttattg tctcatgagc ggatacatat ttgaatgtat ttagaaaaat aaacaaatag
11941 gggttccgcg cacatttccc cgaaaagtgc cac

SEQ ID NO:98 (pTnMCS (CHOVep-prepro-HCPro-CPA))
    1 ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga
   61 ccgctacact tgccagcgcc ctagcgcccg ctcctttcgc tttcttccct tcctttctcg
  121 ccacgttcgc cggcatcaga ttggctattg gccattgcat acgttgtatc catatcataa
  181 tatgtacatt tatattggct catgtccaac attaccgcca tgttgacatt gattattgac
  241 tagttattaa tagtaatcaa ttacgggggtc attagttcat agcccatata tggagttccg
  301 cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt
  361 gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca
  421 atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc
  481 aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta
  541 catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac
  601 catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg
  661 atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg
  721 ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt
```

781 acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg
841 ccatccacgc tgttttgacc tccatagaag acaccgggac cgatccagcc tccgcggccg
901 ggaacggtgc attggaacgc ggattccccg tgccaagagt gacgtaagta ccgcctatag
961 actctatagg cacacccctt tggctcttat gcatgctata ctagtttttg cttgggcct
1021 atacaccccc gcttccttat gctataggtg atggtatagc ttagcctata ggtgtgggtt
1081 attgaccatt attgaccact ccctcattgg tgacgatact ttccattact aatccataac
1141 atggctcttt gccacaacta tctctattgg ctatatgcca atactctgtc cttcagagac
1201 tgacacggac tctgtatttt tacaggatgg ggtcccattt attatttaca aattcacata
1261 tacaacaacg ccgtcccccg tgcccgcagt tttattaaaa catagcgtgg gatctccacg
1321 cgaatctcgg gtacgttcc cggacatggg ctcttctccg gtagcggcgg agcttccaca
1381 tccgagccct ggtcccatgc ctccagcggc tcatggtcgc tcggcagctc cttgctccta
1441 acagtggagg ccagacttag gcacagcaca atgcccacca ccaccagtgt gccgcacaag
1501 gccgtggcgg tagggtatgt gtctgaaaat gagcgtggag attgggctcg cacggctgac
1561 gcagatggaa gacttaaggc agcggcagaa gaaga*gcag gcagctgagt tgttgtattc
1621 tgataagagt cagaggtaac tcccgttgcg gtgccgttaa cggtggaggg cagtgtagtc
1681 tgagcagtac tcgttgctgc cggcgcgcc accagacata atagctgaca gactaacaga
1741 ctgttccttt ccatgggtct tttctgcagt caccgtcgga ccatgtgcga actcgatatt
1801 ttacacgact ctctttacca attctgcccc gaattacact taaaacgact caacagctta
1861 acgttggctt gccacgcatt acttgactgt aaaactctca ctcttaccga acttggccgt
1921 aacctgccaa ccaaagcgag aacaaaacat aacatcaaac gaatcgaccg attgttaggt
1981 aatcgtcacc tccacaaaga gcgactcgct gtataccgtt ggcatgctag ctttatctgt
2041 tcgggcaata cgatgcccat tgtacttgtt gactggtctg atattcgtga gcaaaaacga
2101 cttatggtat tgcgagcttc agtcgcacta cacggtcgtt ctgttactct ttatgagaaa
2161 gcgttcccgc tttcagagca atgttcaaag aaagctcatg accaatttct agccgtcctt
2221 gcgagcattc taccgagtaa caccacaccg ctcattgtca gtgatgctgg ctttaaagtg
2281 ccatggtata aatccgttga gaagctcggt tggtactggt taagtcgagt aagaggaaaa
2341 gtacaatatg cagacctagg agcggaaaac tggaaaccta tcagcaactt acatgatatg
2401 tcatctagtc actcaaagac tttaggctat aagaggctga ctaaaagcaa tccaatctca
2461 tgccaaattc tattgtataa atctcgctct aaaggccgaa aaaatcagcg ctcgacacgg
2521 actcattgtc accacccgtc acctaaaatc tactcagcgt cggcaaagga gccatgggtt
2581 ctagcaacta acttacctgt tgaaattcga acacccaaac aacttgttaa tatctattcg
2641 aagcgaatgc agattgaaga aaccttccga gacttgaaaa gtcctgccta cggactaggc
2701 ctacgccata gccgaacgag cagctcagag cgttttgata tcatgctgct aatcgccctg
2761 atgcttcaac taacatgttg gcttgcgggc gttccatgctc agaaacaagg ttgggacaag
2821 cacttccagg ctaacacagt cagaaatcga aacgtactct caacagttcg cttaggcatg
2881 gaagttttgc ggcattctgg ctacacaata acaagggaag acttactcgt ggctgcaacc
2941 ctactagctc aaaatttatt cacacatggt tacgctttgg ggaaattatg aggggatcgc
3001 tctagagcga tccgggatct cgggaaaagc gttggtgacc aaaggtgcct tttatcatca
3061 ctttaaaaat aaaaaacaat tactcagtgc ctgttataag cagcaattaa ttatgattga

```
3121 tgcctacatc acaacaaaaa ctgatttaac aaatggttgg tctgccttag aaagtatatt
3181 tgaacattat cttgattata ttattgataa taataaaaac cttatcccta tccaagaagt
3241 gatgcctatc attggttgga atgaacttga aaaaaattag ccttgaatac attactggta
3301 aggtaaacgc cattgtcagc aaattgatcc aagagaacca acttaaagct ttcctgacgg
3361 aatgttaatt ctcgttgacc ctgagcactg atgaatcccc taatgatttt ggtaaaaatc
3421 attaagttaa ggtggataca catcttgtca tatgatcccg gtaatgtgag ttagctcact
3481 cattaggcac cccaggcttt acactttatg cttccggctc gtatgttgtg tggaattgtg
3541 agcggataac aatttcacac aggaaacagc tatgaccatg attacgccaa gcgcgcaatt
3601 aaccctcact aaagggaaca aaagctggag ctccaccgcg gtggcggccg ctctagaact
3661 agtggatccc ccgggctgca ggaattcgat atcaagctta tcgataccgc tgacctcgag
3721 ctgcagaaaa atgccaggtg gactatgaac tcacatccaa aggagcttga cctgatacct
3781 gattttcttc aaacagggga aacaacacaa tcccacaaaa cagctcagag agaaaccatc
3841 actgatggct acagcaccaa ggtatgcaat ggcaatccat tcgacattca tctgtgacct
3901 gagcaaaatg atttctctct ccatgaatgg ttgctcttt ccctcatgaa aaggcaattt
3961 ccacactcac aatatgcgac aaagacaaac agagaacaat taatgtgctc cttcctaatg
4021 tcaaaattgt agtggcaaag aggagaacaa aatctcaagt tctgagtagg ttttagtgat
4081 tggataagag gctttgacct gtgagctcac ctggacttca tatccttttg gataaaaagt
4141 gcttttataa ctttcaggtc tccgagtctt tattcatgag actgttggtt tagggacaga
4201 cccacaatga aatgcctggc ataggaaagg gcagcagagc cttagctgac cttttcttgg
4261 gacaagcatt gtcaaacaat gtgtgacaaa actatttgta ctgctttgca cagctgtgct
4321 gggcagggcg atccattgcc acctatccca ggtaaccttc caactgcaag aagattgttg
4381 cttacctcct ctagaaagct tctgcagact gacatgcatt tcataggtag agataacatt
4441 tactgggaag cacatctatc atcacaaaaa gcaggcaaga ttttcagact ttcttagtgg
4501 ctgaaataga agcaaaagac gtaattaaaa acaaaatgaa acaaaaaaaa tcagttgata
4561 cctgtggtgt agacatccag caaaaaaata ttatttgcac taccatcttg tctttagtcc
4621 tcagacttag caaggagaat gtagatttcc acagtatata tgtttttcaca aaaggaagga
4681 gagaaacaaa agaaaatggc actgactaaa cttcagctag tggtatagga aagtaattct
4741 gcttaacaga gattgcagtg atctctatgt atgtcctgaa gaattatgtt gtacttttct
4801 cccccatttt taaatcaaac agtgctttac agaggtcaga atggtttctt tactgtttgt
4861 caattctatt atttcaatac agaacaatag cttctataac tgaaatatat ttgctattgt
4921 atattatgat tgtcccctcga accatgaaca ctcctccagc tgaatttcac aattcctctg
4981 tcatcctgcca ggccattaag ttattcatgg aagatctttg aggaacactg caagttcata
5041 tcataaacac atttgaaatt gagtattggt ttgcattgta tggagctatg tttgctgtga
5101 tcctcagaaa aaaagtttgt tataaagcat tcacacccat aaaagatag atttaaatat
5161 tccagctata ggaaagaaag tgcgtctgct cttcactcta gtcttcagttg gctccttcac
5221 atgcatgctt ctttatttct cctattttgt caagaaaata ataggtcacg tcttgttctc
5281 acttatgtcc tgcctagcat ggcttcagatg cacgttgtag atacaagaag gatcaaatga
5341 aacagacttc tggtctgtta cctacaacca tagtaataag cacactaact aataattgct
5401 aattatgttt tccatcttta aggttcccat atttttctgt ttttttaaag atcccattat
```

```
5461 ctggttgtaa ctgaagctca atggaacatg agcaatattt cccagtcttc tctcccatcc
5521 aacagtcctg atggattagc agaacaggca gaaaacacat tgttacccag aattaaaaac
5581 taatatttgc tctccattca atccaaaatg gacctattga aactaaaatc taacccaatc
5641 ccattaaatg atttctatgg tgtcaaaggt caaacttctg aagggaacct gtgggtgggt
5701 cacaattcag gctatatatt ccccagggct cagccagtgg atcaatcatt catctcgtga
5761 ctttctcgtg tgtggtgttt acctatatat ctaaatttaa tatttcgttt attaaaattt
5821 aatatatttc gacgatgaat ttcttcaagga tatttttctt cgtgttcgct ttggttctgg
5881 ctttgtcaac agtttcggct gcgccagagc cgaaaggtac ccaggtgcag ctgcaggagt
5941 cgggggggag cttgtaaaag ccggggggct cccttagagt ctcctgtgca gcctctggat
6001 tcactttcag aaacgcctgg atgagctggg tccgccaggc tccagggaag gggctggagt
6061 gggtcggccg tattaaaagc aaaattgatg gtgggacaac agactatgct gcacccgtga
6121 aaggcagatt caccatctca agagatgatt caaaaaacac gttatatctg caaatgaata
6181 gcctgaaagc cgaggacaca gccgtatatt actgtaccac gggattatg ataacatttg
6241 ggggagttat ccctcccccg aattgggggcc agggaccccct ggtcacgtcc tcctcagcct
6301 ccaccaaggg cccatcggtc ttccccctgg caccctcctc caagagcacc tctgggggca
6361 cagcggccct gggctgcctg gtcaaggact acttccccga acggtgacg gtgtcgtgga
6421 actcaggcgc cctgaccagc ggcgtgcaca cctttccggc tgtcctacag tcctcaggac
6481 tctacttcct tagcaacgtg gtgaccgtgc cctccagcag cttgggcacc cagacctaca
6541 tctgcaacgt gaatcacaag cccagcaaca ccaaggtgga caagaaagtt gagcccaaat
6601 cttgtgacaa aactcacaca tgcccaccgt gcccagcacc tgaactcctg gggggaccgt
6661 cagtcttcct ctcccccca aaacccaagg acaccctcat gatctcccgg accctgagg
6721 tcacatgcgt ggtggtggac gtgagccacg aagaccctga ggtcaagttc aactggtacg
6781 tggacggcgt ggaggtgcat aatgccaaga caaagccgcg ggaggagcag tacaacagca
6841 cgtaccgtgt ggtcagcgtc ctcaccgtcc tgcaccagga ctggctgaat ggcaaygagt
6901 acaagtgcaa ggtctccaac aaagccctcc cagccccat cgagaaaacc atctccaaag
6961 ccaaagggca gccccgagaa ccacaggtgt acaccctgcc cccatcccgg gatgagctga
7021 ccaagaacca ggtcagcctg acctgcctgg tcaaaggctt ctatcccagc gacatcgccg
7081 tggagtggga gagcaatggg cagccggaga acaactacaa gaccacgcct cccgtgctgg
7141 actccgacgg ctccttcttc ctctacagca agctcaccgt ggacaagagc aggtggcagc
7201 aggggaacgt cttctcatgc tccgtgatgc atgaggctct gcacaaccac tacacgcaga
7261 agagcctctc cctgtctccg ggtaaagcgc cagagccgaa gctttcctat gagctgacac
7321 agccacccctc ggtgtcagtg tccccaggac aaacggccag gatcacctgc tctggagatg
7381 catgccaga aaaatatgtt tattggtacc agcagaagtc aggccaggcc cctgtggtgg
7441 tcatctatga ggacagcaaa cgaccctccg ggatccctga gagttctct ggctccagct
7501 cagggacaat ggccaccttg actatcagtg gggcccaggt ggaagatgaa ggtgactact
7561 actgttactc aactgacagc agtggttatc atagggaggt gttcagcgga gggaccaagc
7621 tgaccgtcct aggtcagccc aaggctgccc cctcggtcac tctgttccca ccctcctctg
7681 aggagcttca agccaacaag gccacactgg tgtgtctcat aagtgactcc taccgggag
7741 ccgtgacagt ggcctggaag gcagatagca gccccgtcaa ggcgggagtg gagaccacca
```

7801 caccctccaa acaaagcaac aacaagtacg cggccagcag ctacctgagc ctgacgcttg
7861 agcagtggaa gtcccacaaa agctacagct gccaggtcac gcatgaaggg agcaccgtgg
7921 agaagacagt ggccctgca gaatgttcac cgcggagga gggaagggcc cttttgaag
7981 ggggaggaaa cttcgcgcca tgactccttct cgtgcccccc gcacggaaca ctgatgtgca
8041 gagggccctc tgccatgct gcttcctctg ccctcctcg tcacctctgaa tgtggcttct
8101 ttgctactgc cacagcaaga aataaaatct caacatctaa atgggtttcc tgagattttt
8161 caagagtcgt taagcacatt ccttccccag caccccttgc tgcaggccag tgccaggcac
8221 caacttggct actgctgccc atgagagaaa tccagttcaa tattttccaa agcaaaatgg
8281 attacatatg ccctagatcc tgattaacag gtgttttgta ttatctgtgc tttcgcttca
8341 cccacattat cccatgcct ccctcgagg ggggcccgg tacccaattc gccctatagt
8401 gagtcgtatt acgcgcgctc actggccgtc gtttacaac gtcgtgactg ggaaaaccct
8461 ggcgttaccc aacttaatcg ccttgcagca catccccctt tcgccagctg gcgtaatagc
8521 gaagaggccc gcaccgatcg ccctttcccaa cagttgcgca gcctgaatgg cgaatggaaa
8581 ttgtaagcgt taatatttg ttaaaattcg cgttaaattc ttgttaaatc agctcatttt
8641 ttaaccaata ggccgaaatc ggcaaaatcc cttacaaatc aaaagaatag accgagatag
8701 ggttgagtgt tgttccagtt tggaacaaga gtccactatt aaagaacgtg gactccaacg
8761 tcaaagggcg aaaaccgtc tatcagggcg atggcccact actccgggat catatgacaa
8821 gatgtgtatc cacttaact taatgatttt taccaaaatc attagggat tcatcagtgc
8881 tcagggtcaa cgagaattaa cattccgtca ggaaagctta tgatgatgat gtgcttaaaa
8941 acttactcaa tggctggtta tgcatatcgc aatacatgcg aaaaacctaa aagagcttgc
9001 cgataaaaaa ggccaattta ttgctattta ccgcggcttt ttattgagct tgaaagataa
9061 ataaaataga taggtttat ttgaagctaa atcttctttta tcgtaaaaaa tgccctcttg
9121 ggttatcaag agggtcatta tatttcgcgg aataacatca ttggtgacg aaataactaa
9181 gcacttgtct cctgtttact cccctgagct tgaggggtta acatgaaggt catcgctagc
9241 aggataataa tacagtaaaa cgctaaacca ataatccaaa tccagccatc ccaacttggt
9301 agtgaatgat tataaataac agcaaacagt aatgggccaa taacaccggt tgcattggta
9361 aggctcacca ataatccctg taaagcacct tgctgatgac tctttgtttg gatagacatc
9421 actccctgta atgcaggtaa agcgatccca ccaccagcca ataaaattaa aacagggaaa
9481 actaaccaac cttcagatat aaacgctaaa aaggcaaatg cactactatc tgcaataaat
9541 ccgagcagta ctgccgtttt ttcgcccatt tagtggctat tcttcctgcc acaaaggctt
9601 ggaatactga gtgtaaaaga ccaagacccg taatgaaaag ccaaccatca tgctattcat
9661 catcacgatt tctgtaatag caccacaccg tgctggattg gctatcaatg cgctgaaata
9721 ataatcaaca aatggcatcg ttaaataagt gatgtatacc gatcagcttt tgttcccttt
9781 agtgagggtt aatgcgcgc ttggcgtaat catggtcata gctgtttcct gtgtgaaatt
9841 gttatccgct cacaattcca cacaacatac gagccggaag cataaagtgt aaagcctggg
9901 gtgcctaatg agtgagctaa ctcacattaa ttgcgttgcg ctcactgccc gctttccagt
9961 cgggaaacct gtcgtgccag ctgcattaat gaatcggcca acgcgcgggg agaggcggtt
10021 tgcgtattgg gcgctcttcc gcttcctcgc tcactgactc gctgcgctcg gtcgttcggc
10081 tgcggcgagc ggtatcagct cactcaaagg cggtaatacg gttatccaca gaatcagggg

```
10141 ataacgcagg aaagaacatg tgagcaaaag gccagcaaaa ggccaggaac cgtaaaaagg
10201 ccgcgttgct ggcgtttttc cataggctcc gcccccctga cgagcatcac aaaaatcgac
10261 gctcaagtca gaggtggcga aacccgacag gactataaag ataccaggcg tttccccctg
10321 gaagctccct cgtgcgctct cctgttccga ccctgccgct taccggatac ctgtccgcct
10381 ttctcccttc gggaagcgtg gcgctttctc atagctcacg ctgtaggtat ctcagttcgg
10441 tgtaggtcgt tcgctccaag ctgggctgtg tgcacgaacc cccgttcagc ccgaccgct
10501 gcgccttatc cggtaactat cgtcttgagt ccaacccggt aagacacgac ttatcgccac
10561 tggcagcagc cactggtaac aggattagca gagcgaggta tgtaggcggt gctacagagt
10621 tcttgaagtg gtggcctaac tacggctaca ctagaaggac agtatttggt atctgcgctc
10681 tgctgaagcc agttaccttc ggaaaaagag ttggtagctc ttgatccggc aaacaaacca
10741 ccgctggtag cggtggtttt ttgtttgca agcagcagat tacgcgcaga aaaaaggat
10801 ctcaagaaga tccttttgatc ttttctacgg ggtctgacgc tcagtggaac gaaaactcac
10861 gttaagggat ttggtcatg agattatcaa aaggatctt cacctagatc ctttaaatt
10921 aaaaatgaag tttaaatca atctaaagta tatatgagta aacttggtct gacagttacc
10981 aatgcttaat cagtgaggca cctatctcag cgatctgtct atttcgttca tccatagttg
11041 cctgactccc cgtcgtgtag ataactacga tacgggaggg cttaccatct ggccccagtg
11101 ctgcaatgat accgcgagac ccacgctcac cggctccaga tttatcagca ataaaccagc
11161 cagccggaag ggccgagcgc agaagtggtc ctgcaacttt atccgcctcc atccagtcta
11221 ttaattgttg ccgggaagct agagtaagta gttcgccagt taatagtttg cgcaacgttg
11281 ttgccattgc tacaggcatc gtggtgtcac gctcgtcgtt tggtatggct tcattcagct
11341 ccggttccca acgatcaagg cgagttacat gatcccccat gttgtgcaaa aaagcggtta
11401 gctccttcgg tcctccgatc gttgtcagaa gtaagttggc cgcagtgtta tcactccatgg
11461 ttatggcagc actgcataat tctcttactg tcatgccatc cgtaagatgc ttttctgtga
11521 ctggtgagta ctcaaccaag tcattctgag aatagtgtat gcggcgaccg agttgctctt
11581 gcccggcgtc aatacgggat aataccgcgc cacatagcag aactttaaaa gtgctcatca
11641 ttggaaaacg ttcttcgggg cgaaaactct caaggatctt accgctgttg agatccagtt
11701 cgatgtaacc cactcgtgca cccaactgat cttcagcatc tttactttc accagcgttt
11761 ctgggtgagc aaaaacagga aggcaaaatg ccgcaaaaaa gggaataagg gcgacacgga
11821 aatgttgaat actcatactc ttccttttc aatattattg aagcatttat cagggttatt
11881 gtctcatgag cggatacata ttgaatgta ttagaaaaa taaacaaata ggggttccgc
11941 gcacatttcc ccgaaaagtg ccac
//
SEQ ID NO:99 (pTnMCS(CHOvep-prepro-HCPro-Lys-CPA))
   1 ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga
  61 ccgctacact tgccagcgcc ctagcgcccg ctcctttcgc tttcttccct tcctttctcg
 121 ccacgttcgc cggcatcaga ttggctattg gccattgcat acgttgtatc catatcataa
 181 tatgtacatt tatattggct catgtccaac attaccgcca tgttgacatt gattattgac
 241 tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg
```

```
301 cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt
361 gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca
421 atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc
481 aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta
541 catgaccttа tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac
601 catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg
661 atttccaagt ctcacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg
721 ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt
781 acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg
841 ccatccacgc tgtttgacct tccatagaag acaccgggac cgatccagcc tccgcggccg
901 ggaacggtgc attggaacgc ggattccccg tgccaagagt gacgtaagta ccgcctatag
961 actctatagg cacacccctt tggctcttat gcatgctata ctgttttgg cttggggcct
1021 atacacccc gcttccttat gctataggtg atggtatagc ttagcctata ggtgtgggtt
1081 attgaccatt atgaccact ccctcattgg tgacgatact ttccattact aatccataac
1141 atggctcttt gccacaacta tctctattgg ctatacgcca atactctgtc cttcagagac
1201 tgacacggac tctgtatttt tacaggatgg ggtcccattt attatttaca aattcacata
1261 tacaacaacg ccgtcccccg tgcccgcagt tttattaaaa catagcgtgg gatctccacg
1321 cgaatctcgg gtacgtgttc cggacatggg ctcttctccg gtagcggcgg agcttccaca
1381 tccgagccct ggtcccatgc ctccagcggc tcatggtcgc tcggcagctc cttgctccta
1441 acagtggagg ccagacttag gcacagcaca atgcccacca ccaccagtgt gccgcacaag
1501 gccgtggcgg tagggtatgt gtctgaaaat gagcgtggag attgggctcg cacggctgac
1561 gcagatggaa gacttaaggc agcggcagaa gaagatgcag gcagctgagt tgttgtattc
1621 tgataagagt cagaggtaac tcccgttgcg gtgctgttaa cggtggaggg cagtgtagtc
1681 tgagcagtac tcgttgctgc cgcgcgccc accagacata atagctgaca gactaacaga
1741 ctgttccttt ccatgggtct tttctgcagt caccgtcgga ccatgtgcga actcgtatt
1801 ttacacgact ctctttacca attctgcccc gaattacact taaaacgact caacagctta
1861 acgttggctt gccacgcatt acttgactgt aaaactcttca ctcttaccga acttggccgt
1921 aacctgccaa ccaaagcgag aacaaaacat aacatcaaac gaatcgaccg attgttaggt
1981 aatcgtcacc tccacaaaga gcgactcgct gtataccgtt ggcatgctag ctttatctgt
2041 tcgggcaata cgatgcccat tgtacttgtt gactggtctg atattcgtga gcaaaaacga
2101 cttatggtat tgcgagcttc agtcgcacta cacggtcgtt ctgttactct ttatgagaaa
2161 gcgttcccgc tttcagagca atgttcaaag aaagcttcatg accaatttct agccgacctt
2221 gcgagcattc taccgagtaa caccacaccg ctcattgtca gtgatgctgg ctttaaagtg
2281 ccatggtata aatccgttga gaagctgggt tggtactggt taagtcgagt aagaggaaaa
2341 gtacaatatg cagacttagg agcggaaaac tggaaaccta tcagcaactt acatgatatg
2401 tcatctagtc actcaaagac tttaggctat aagaggctga ctaaaagcaa tccaatctca
2461 tgccaaattc tattgtataa atctcgctct aaaggccgaa aaaatcagcg ctcgacacgg
2521 actcattgtc accacccgtc acctaaaatc tactcagcgt cggcaaagga gccatgggtt
2581 ctagcaacta acttacctgt tgaaattcga acacccaaac aacttgttaa tatcttatcg
```

```
2641 aagcgaatgc agattgaaga aaccttccga gacttgaaaa gtcctgccta cggactaggc
2701 ctacgccata gccgaacgag cagctcagag cgttttgata tcatgctgct aatcgccctg
2761 atgcttcaac taacatgttg gcttgcgggc gttcatgctc agaaacaagg ttgggacaag
2821 cacttccagg ctaacacagt cagaaatcga aacgtactct caacagttcg cttaggcatg
2881 gaagttttgc ggcattctgg ctacacaata acaagggaag acttactcgt ggctgcaacc
2941 ctactagctc aaaatttatt cacacatggt tacgctttgg ggaaattatg aggggatcgc
3001 tctagagcga tccgggaaagc gttggtgacc aaaggtgcct tttatcatca
3061 ctttaaaaat aaaaaacaat tactcagtgc ctgttataag cagcaattaa ttatgattga
3121 tgcctacatc acaacaaaaa ctgatttaac aaatggttgg tctgccttag aaagtatatt
3181 tgaacattat cttgattata ttattgataa taataaaaac cttatcccta tccaagaagt
3241 gatgcctatc attggttgga atgaacttga aaaaaattag ccttgaatac attactggta
3301 aggtaaacgc cattgtcagc aaattgatcc aagagaacca acttaaagct ttcctgacgg
3361 aatgttaatt ctcgttgacc ctgagcactg atgaatcccc taatgatttt ggtaaaaatc
3421 attaagttaa ggtggataca catcttgtca tatgat-ccg gtaatgtgag ttagctcact
3481 cattaggcac cccaggcttt acactttatg cttccggctc gtatgttgtg tggaattgtg
3541 agcggataac aatttcacac aggaaacagc tatgaccatg attacgccaa gcgcgcaatt
3601 aaccctcact aaagggaaca aaagctggag ctccaccgcg gtggcggccg ctctagaact
3661 agtggatccc cgggctgca ggaattcgat atcaagctta tcgataccgc tgacctcgag
3721 ctgcagaaaa atgccaggtg gactatgaac tcacatccaa aggagcttga cctgatacct
3781 gatttcttc aaacagggga aacaacacaa tcccacaaaa cagctcagag agaaaccatc
3841 actgatggct acagcaccaa ggtatgcaat ggcaatccat tcgacattca tctgtgacct
3901 gagcaaaatg atttctctct ccatgaatgg ttgcttcttt ccctcatgaa aaggcaattt
3961 ccacactcac aatatgcgac aaagacaaac agagaacaat taatgtgctc cttcctaatg
4021 tcaaaattgt agtggcaaag aggagaacaa aatctcaagt tctgagtagg tttagtgat
4081 tggataagag gctttgacct gtgagctcac ctggacttca tatcctttg gataaaaagt
4141 gcttttataa ctttcaggtc tccgagtctt tattcatgag actgttggtt tagggacaga
4201 cccacaatga aatgcctggc ataggaaagg gcagcagagc cttagctgac cttttcttgg
4261 gacaagcatt gtcaaacaat gtgtgacaaa actatttgta ctgcttgca cagctgtgct
4321 gggcagggcg atccattgcc acctatccca ggtaaccttc caactgcaag aagattgttg
4381 cttactctct ctagaaagct tctgcagact gacatgcatt tcataggtag agataacatt
4441 tactgggaag cacatctatc atcacaaaaa gcaggcaaga ttttcagact ttcttagtgg
4501 ctgaaataga agcaaaagac gtaattaaaa acaaaaatga acaaaaaaaa tcagttgata
4561 cctgtggtgt agacatccag caaaaaaata ttatttgcac taccatcttg tcttaagtcc
4621 tcagacttag caaggagaat gtagatttcc acagtatata tgttttcaca aaaggaagga
4681 gagaaacaaa agaaaatggc actgactaaa cttcagctag tggtataagga aagtaattct
4741 gcttaacaga gattgcagtg atctctatgt atgtcctgaa gaattatgtt gtacttttt
4801 ccccatttt taaatcaaac agtgctttac agaggtcaga atggtttctt tactgtttgt
4861 caattctatt atttcaatac agaacaatag cttctataac tgaaatatat ttgctattgt
4921 atattatgat tgtcctcga accatgaaca ctcctccagc tgaattccac aattcctctg
```

```
4981 tcatctgcca ggccattaag ttattcatgg aagatctttg aggaacactg caagttcata
5041 tcataaacac attgaaatt gagtattggt ttgcattgta tggagctatg tttgctgta
5101 tcctcagaaa aaagtttgt tataaagcat tcacacccat aaaaagatag atttaaatat
5161 tccagctata ggaaagaaag tgcgtcttgct cttcactcta gtcttcagttg gctccttcac
5221 atgcatgctt ctttatttct cctatttttgt caagaaaata ataggttcacg tcttgttctc
5281 acttatgtcc tgcctagcat ggctcagatg cacgttgtag atacaagaag gatcaaatga
5341 aacagacttc tggtctgtta cctacaacca tagtaataag cacactaact aataattgct
5401 aattatgttt tccatctcta aggttcccat attttctgt ttcttaaag atcccattat
5461 ctggttgtaa ctgaagctca atggaacatg agcaatattt cccagtcttc tctcccatcc
5521 aacagtcctg atggattagc agaacaggca gaaacacat tgttacccag aattaaaaac
5581 taatatttgc tctccattca atccaaaatg gacctattga aactaaaatc taacccaatc
5641 ccattaaatg attctatgg tgtcaaaggt caaacttctg aagggaacct gtgggtgggt
5701 cacaattcag gcttatatatt ccccagggct cagccagtgg atcaatcatt catctcgtga
5761 ctcctcgtg tgtggtgttt acctatatat ctaaatttaa tatttcgttt attaaaattt
5821 aatatatttc gacgatgaat ttctcaagga tatttctctt cgtgttcgct ttggttcttgg
5881 ctttgtcaac agtttcggct gcgccagagc cgaaaggtac ccaggtgcag ctgcaggagt
5941 cgggggggagg ctggtaaag ccgggggggt cccttagagt ctcctgtgca gcctctggat
6001 tcactttcag aaacgcctgg atgagctggg tccgccaggc tccagggaag gggctggagt
6061 gggtcggccg tattaaaagc aaaattgatg gtgggacaac agactatgct gcacccgtga
6121 aaggcagatt caccatctca agagatgatt caaaaaacac gttatatctg caaatgaata
6181 gcctgaaagc cgaggacaca gccgtatatt actgtaccac ggggattatg ataacatttg
6241 ggggagttat ccctcccccg aattgggccc agggaaccct ggtcaccgtc tcctcagcct
6301 ccaccaaggg cccatcggtc ttccccctgg caccctcctc caagagcacc tctgggggca
6361 cagcggccct gggctgcctg gtcaaggact acttccccga accggtgacg gtgtcgtgga
6421 actcaggcgc cctgaccagc ggcgtgcaca cctttccggc tgtcctacag tcctcaggac
6481 tctacttcct tagcaacgtg gtgaccgtgc cctccagcag ctgggcacc cagacctaca
6541 tctgcaacgt gaatcacaag cccagcaaca ccaaggtgga caagaaagtt gagcccaaat
6601 cttgtgacaa aactcacaca tgcccaccgt gcccagcacc tgaactcctg gggggaccgt
6661 cagtcttcct cttccccca aaacccaagg acaccctcat gatctcccgg acccctgagg
6721 tcacatgcgt ggtggtggac gtgagccacg aagaccctga ggtcaagttc aactggtacg
6781 tggacggcgt ggaggtgcat aatgccaaga caaagccgcg ggaggagcag tacaacagca
6841 cgtaccgtgt ggtcagcgtc ctcaccgtcc tgcaccagga ctggctgaat ggcaaggagt
6901 acaagtgcaa ggtctccaac aaagccctcc cagccccat cgagaaaacc atctccaaag
6961 ccaaagggca gccccgagaa ccacaggtgt acaccctgcc cccatcccgg gatgagctga
7021 ccaagaacca ggtcagcctg acctgcctgg tcaaaggctt ctatcccagc gacatcgccg
7081 tggagtggga gagcaatggg cagccggaga acaactacaa gaccacgcct cccgtgctgg
7141 actccgacgg ctccttcttc ctctacagca agctcaccgt ggacaagagc aggtggcagc
7201 aggggaacgt cttctcatgc tccgtgatgc atgaggctct gcacaaccac tacacgcaga
7261 agagcctctc cctgtctccg ggtaaagcgc cagagccgaa aaagctttcc tatgagctga
```

```
7321 cacagccacc ctcggtgtca gtgtccccag gacaaacggc caggatcacc tgctctggag
7381 atgcattgcc agaaaaatat gtttattggt accagcagaa gtcaggccag gcccctgtgg
7441 tggtcatcta tgaggacagc aaacgacgct ccgggatccc tgaggagattc tctggctcca
7501 gctcagggac aatggccacc ttgactatca gtgggccca ggtggaagat gaaggtgact
7561 actactgtta ctcaactgac agcagtggtt atcataggga ggtgttcagc ggagggacca
7621 agctgaccgt cctaggtcag cccaaggctg cccctcggt cactctgttc ccaccctcct
7681 ctgaggagct tcaagccaac aaggccacac tggtgtgtct cataagtgac tcctacccgg
7741 gagccgtgac agtggcctgg aaggcagata gcagccccgt caaggcggga gtggagacca
7801 ccacaccctc caaacaaagc aacaacaagt acgcggccag cagctacctg agcctgacgc
7861 ttgagcagtg gaagtcccac aaaagctaca gctgccaggt cacgcatgaa gggagcaccg
7921 tggagaagac agtggcccct gcagaatgtt caccgcggag ggagggaagg gccctttttg
7981 aaggggggagg aaacttcgcg ccatgactcc tctcgtgccc ccgcacgga acactgatgt
8041 gcagagggcc ctctgccatt gctgcttcct ctgccttcc tcgtcactct gaatgtggct
8101 tctttgctac tgccacagca agaaataaaa tctcaaraatc taaatgggtt tcctgagatt
8161 ttcaagagt cgttaagcac attccttccc cagcaccct tgctgcaggc cagtgccagg
8221 caccaacttg gctactgctg cccatgagag aaatccagtt caatatttc caaagcaaaa
8281 tggattacat atgccctaga tcctgattaa caggtgtttt gtattatctg tgctttcgct
8341 tcacccacat tatcccattg cctcccctcg agggggggcc cggtacccaa ttcgccctat
8401 agtgagtcgt attacgcgcg ctcactggcc gtcgttttac aacgtcgtga ctgggaaaac
8461 cctggcgtta cccaacttaa tcgccttgca gcacatcccc ctttcgccag ctggcgtaat
8521 agcgaagagg cccgcaccga tcgcccttcc caacagttgc gcagcctgaa tggcgaatgg
8581 aaattgtaag cgttaatatt ttgttaaaat tcgcgttaaa ttttgttaa atcagctcat
8641 ttttaacca ataggccgaa atcggcaaaa tcccttataa atcaaaagaa tagaccgaga
8701 tagggttgag tgttgttcca gtttggaaca agagtccact attaaagaac gtggactcca
8761 acgtcaaagg gcgaaaaacc gtctatcagg gcgatggccc actactcggg gatcatatga
8821 caagatgtgt atccacctta acttaatgat ttttaccaaa atcattaggg gattcatcag
8881 tgctcagggt caacgagaat taacattccg tcaggaaagc ttatgatgat gatgtgctta
8941 aaaacttact caatggctgg ttatgcatat cgcaatacat gcgaaaaacc taaaagagct
9001 tgccgataaa aaaggccaat ttattgctat ttaccgcggc ttttattatga gcttgaaaga
9061 taaataaaat agataggttt tatttgaagc taaatcttct ttatcgtaaa aaatgcgctc
9121 ttgggttatc aagagggtca ttatatttcg cggaataaca tcatttggtg acgaaataac
9181 taagcacttg tctcctgttt actcccctga gcttgagggg ttaacatgaa ggtcatcgat
9241 agcaggataa taatacagta aaacgctaaa ccaataatcc aaatccagcc atcccaaatt
9301 ggtagtgaat gattataaat aacagcaaac agtaatgggc caataacacc ggttgcattg
9361 gtaaggctca ccaataatcc ctgtaaagca ccttgctgat gactttttgt ttggatagac
9421 atcactcccct gtaatgcagg taaagcgatc ccaccaccag ccaataaaat taaaacaggg
9481 aaaactaacc aaccttcaga tataaacgct aaaaaggcaa atgcactact atctgcaata
9541 aatccgagca gtactgccgt ttttcgccc atttagtggc tatttctcct gccacaaagg
9601 cttggaatac tgagtgtaaa agaccaagac ccgtaatgaa aagccaacca tcatgctatt
```

```
 9661 catcatcacg atttctgtaa tagcaccaca ccgtgctgga ttggctatca atgcgctgaa
 9721 ataataatca acaaatggca tcgttaaata agtatgtat accgatcagc tttgttccc
 9781 tttagtgagg gttaattgcg cgcttggcgt aatcatggtc atagctgttt cctgtgtgaa
 9841 attgttatcc gctcacaatt ccacacaaca tacgagccgg aagcataaag tgtaaagcct
 9901 ggggtgccta atgagtgagc taactcacat taattgcgtt gcgctcactg ccgctttcc
 9961 agtcgggaaa cctgtcgtgc cagctgcatt aatgaatcgg ccaacgcgcg gggagaggcg
10021 gtttgcgtat tgggcgctct tccgcttcct cgctcactga ctcgctgcgc tcggtcgttc
10081 ggctgcggcg agcggtatca gctcactcaa aggcggtaat acggttatcc acagaatcag
10141 gggataacgc aggaaagaac atgtgagcaa aaggccagca aaaggccagg aaccgtaaaa
10201 aggccgcgtt gctggcgttt ttccataggc tccgcccccc tgacgagcat cacaaaaatc
10261 gacgctcaag tcagaggtgg cgaaacccga caggactata aagataccag gcgtttcccc
10321 ctggaagctc cctcgtgcgc tctcctgttc cgaccctgcc gcttacggga tacctgtccg
10381 cctttctccc ttcgggaagc gtggcgcttt ctcatagctc acgctgtagg tatctcagtt
10441 cggtgtaggt cgttcgctcc aagctgggct gtgtgcacga acccccgtt cagcccgacc
10501 gctgcgcctt atccggtaac tatcgtcttg agtccaaccc ggtaagacac gacttatcgc
10561 cactggcagc agccactggt aacaggatta gcagagcgag gtatgtaggc ggtgctacag
10621 agttcttgaa gtggtggcct aactacggct acactagaag gacagtattt ggtatctgcg
10681 ctctgctgaa gccagttacc ttcggaaaaa gagttggtag ctcttgatcc ggcaaacaaa
10741 ccaccgctgg tagcggtggt tttttgttt gcaagcagca gattacgcgc agaaaaaaag
10801 gatctcaaga agatcctttg atcttttcta cggggtctga cgctcagtgg aacgaaaact
10861 cacgttaagg gattttggtc atgagattat caaaaaggat cttcacctag atcctttaa
10921 attaaaaatg aagttttaaa tcaatctaaa gtatatatga gtaaacttgg tctgacagtt
10981 accaatgctt aatcagtgag gcacctatct cagcgatctg tctatttcgt tcatccatag
11041 ttgcctgact ccccgtcgtg tagataacta cgatacggga gggcttacca tctggccccа
11101 gtgctgcaat gataccgcga gacccacgct caccggctcc agatttatca gcaataaacc
11161 agccagccgg aagggccgag cgcagaagtg gtcctgcaac tttatccgcc tccatccagt
11221 ctattaattg ttgccgggaa gctagagtaa gtagttcgcc agttaatagt ttgcgcaacg
11281 ttgttgccat tgctacaggc atcgtggtgt cacgctcgtc gtttggtatg gcttcattca
11341 gctccggttc ccaacgatca aggcgagtta catgatcccc catgttgtgc aaaaaagcgg
11401 ttagctcctt cggtcctccg atcgttgtca gaagtaagtt ggccgcagtg ttatcactca
11461 tggttatggc agcactgcat aattctctta ctgtcatgcc atccgtaaga tgcttttctg
11521 tgactggtga gtactcaacc aagtcattct gagaatagtg tatgcggcga ccgagttgct
11581 cttgcccggc gtcaatacgg gataataccg cgccacatag cagaacttta aaagtgctca
11641 tcattggaaa acgttcttcg gggcgaaaac tctcaaggat cttaccgctg ttgagatcca
11701 gttcgatgta acccactcgt gcacccaact gatcttcagc atcttttact ttcaccagcg
11761 tttctgggtg agcaaaaaca ggaaggcaaa atgccgcaaa aaagggaata agggcgacac
11821 ggaaatgttg aatactcata ctcttccttt ttcaatatta ttgaagcatt tatcagggtt
11881 attgtctcat gagcggatac atatttgaat gtatttagaa aataaacaa atagggttc
11941 cgcgcacatt tccccgaaaa gtgccac
```

SEQ ID NO:100 (pTnMCS (CMV-prepro-HCPro-CPA))

```
   1 ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga
  61 ccgctacact tgccagcgcc ctagcgcccg ctcctttcgc tttcttcccc tcctttctcg
 121 ccacgttcgc cggcatcaga ttggctattg gccattgcat acgttgtatc catatcataa
 181 tatgtacatt tatattggct catgtccaac attaccgcca tgttgacatt gattattgac
 241 tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg
 301 cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt
 361 gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca
 421 atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc
 481 aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta
 541 catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac
 601 catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg
 661 atttccaagt ctccacccca ttgacgtcaa tgggactttg tttggcaccc aaaatcaacg
 721 ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt
 781 acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg
 841 ccatccacgc tgttttgacc tccatagaag acaccgggac cgatccagcc tccgcggccg
 901 ggaacggtgc attggaacgc ggattccccg tgccaagagt gacgtaagta ccgcctatag
 961 actctatagg cacacccctt tggctcttat gcatgctata ctgttttttg ctggggcctt
1021 atacacccgc gcttccttat gctataggtg atggtatagc ttagcctata ggtgtgggtt
1081 attgaccatt atgaccact cccctattgg tgacgatact ttccattact aatccataac
1141 atggctcttt gccacaacta tctctattgg ctatatgcca atactctgtc cttcagagac
1201 tgacacggac tctgtatttt tacaggatgg ggtcccattt attatttaca aattcacata
1261 tacaacaacg ccgtcccccg tgcccgcagt tttattaaaa catagcgtgg gatctrcacg
1321 cgaatctcgg gtacgttcc cggacatggg ctcttctccg gtagcggcgg agcttccaca
1381 tccgagccct ggtcccatgc ctccagcggc tcatggtcgc tcggcagctc cttgctccta
1441 acagtggagg ccagacttag gcacagcaca atgcccacca ccaccagtgt gccgcacaag
1501 gccgtggcgg taggtatgt gtctgaaaat gagcgtggag atgggctcg cacggctgac
1561 gcagatggaa gacttaaggc agcggcagaa gaagatgcag gcagctgagt tgttgtattc
1621 tgataagagt cagaggtaac tcccgttgcg gtgctgttaa cggtggaggg cagtgtagtc
1681 tgagcagtac tcgttgctgc cgcgcgccgc accagacata atagctgaca gactaacaga
1741 ctgttccttt ccatgggtct ttctgcagt caccgtcgga ccatgtgcga actcgatatt
1801 ttacacgact ctctttacca attctgcccc gaattacact taaaacgact caacagctta
1861 acgttggctt gccacgcatt acttgactgt aaaactctca ctcttaccga acttggccgt
1921 aacctgccaa ccaaagcgag aacaaaacat aacatcaaac gaatcgaccg attgttaggt
1981 aatcgtcacc tccacaaaga gcgactcgct gtataccgtt ggcatgctag ctttatctgt
2041 tcgggcaata cgatgcccat tgtacttgtt gactggtctg atattcgtga gcaaaaacga
2101 cttatggtat tgcgagcttc agtcgcacta cacggtcgtt ctgttactct ttatgagaaa
2161 gcgttccccgc tttcagagca atgttcaaag aaagctcatg accaatttct agccgacctt
```

```
2221 gcgagcattc taccgagtaa caccacaccg ctcattgtca gtgatgctgg ctttaaagtg
2281 ccatggtata aatccgttga gaagctgggt tggtactggt taagtcgagt aagaggaaaa
2341 gtacaatatg cagacctagg agcggaaaac tggaaaccta tcagcaactt acatgatatg
2401 tcatctagtc actcaaagac tttaggctat aagaggctga ctaaaagcaa tccaatctca
2461 tgccaaattc tattgtataa atctcgctct aaaggccgaa aaaatcagcg ctcgacacgg
2521 actcattgtc accacccgtc acctaaaatc tactcagcgt cggcaaagga gccatgggtt
2581 ctagcaacta acttacctgt tgaaattcga acacccaaac aacttgttaa tatcttattcg
2641 agcgaatgc agattgaaga aaccttccga gacttgaaaa gtcctgccta cggactaggc
2701 ctacgccata gccgaacgag cagctcagag cgttttgata tcatgctgct aatcgccctg
2761 atgcttcaac taacatgttg gcttcgcggc gttcatgctc agaaacaagg ttgggacaag
2821 cacttccagg ctaacacagt cagaaatcga aacgtactct caacagttcg cttaggcatg
2881 gaagttttgc ggcattctgg ctacacaata acaagggaag acttactcgt ggctgcaacc
2941 ctactagctc aaaatttatt cacacatggt tacgctttgg ggaaattatg agggatcgtc
3001 tctagagcga tccgggatct cgggaaaagc gttggtgacc aaaggtgcct tttatcatca
3061 ctttaaaaat aaaaaacaat tactcagtgc ctgttataag cagcaattaa ttatgattga
3121 tgcctacatc acaacaaaaa ctgatttaac aaatggttgg tctgccttag aaagtatatt
3181 tgaacattat cttgattata ttattgataa taataaaaac cttatcccta tccaagaagt
3241 gatgcctatc attggttgga atgaacttga aaaaaattag ccttgaatac attactggta
3301 aggtaaacgc cattgtcagc aaattgatcc aagagaacca acttaaagct ttcctgacgg
3361 aatgttaatt ctcgttgacc ctgagcactg atgaatcccc taatgatttt ggtaaaaatc
3421 attaagttaa ggtggataca catctttgtca tatgatcccg gtaatgtgag ttagctcact
3481 cattaggcac cccaggcttt acactttatg cttccggctc gtatgttgtg tggaattgtg
3541 agcggataac aatttcacac aggaaacagc tatgaccatg attacgccaa gcgcgcaatt
3601 aaccctcact aaagggaaca aaagctggag ctccaccgcg gtggcggccg ctctagaact
3661 agtggatccc ccgggctgca ggaattcgat atcaagctta tcgataccgc tgacctcgag
3721 catcagattg gctattggcc attgcatacg ttgtatccat atcataatat gtacatttat
3781 attggctcat gtccaacatt accgccatgt tgacattgat tattgactag ttattaatag
3841 taatcaatta cggggtcatt agttcatagc ccatatatgg agttccgcgt tacataactt
3901 acggtaaatg gcccgcctgg ctgaccgccc aacgacccgc gcccattgac gtcaataatg
3961 acgtatgttc ccatagtaac gccaatagggg actttccatt gacgtcaatg ggtggagtat
4021 ttacggtaaa ctgcccactt ggcagtacat caagtgtatc atatgtcaag tacgccccct
4081 attgacgtca atgacggtaa atggcccgcc tggcattatg cccagtacat gaccttatgg
4141 gactttccta cttggcagta catctacgta ttagtcatcg ctattaccat ggtgatgcgg
4201 ttttggcagt acatcaatgg gcgtggatag cggtttgact cacggggatt tccaagtctt
4261 cacccattg acgtcaatgg gagtttgttt tggcaccaaa atcaacggga ctttccaaaa
4321 tgtcgtaaca actccgcccc attgacgcaa atgggcggta ggcgtgtacg gtgggaggtc
4381 tatataagca gagctcgttt agtgaaccgt cagatcgcct ggagacgcca tccacgctgt
4441 ttgacctcc atagaagaca ccgggaccga tccagcctcc gcggccggga acggtgcatt
4501 ggaacgcgga tccccgtgc caagagtgac gtaagtaccg cctatagact ctataggcac
```

4561 acccctttgg ctcttatgca tgctatactg ttttggctt ggggcctata cacccccgct
4621 tcctatgct ataggtgatg gtatagctta gcctataggt gtgggttatt gaccattatt
4681 gaccactccc ctattggtga cgatactttc cattactaat ccataacatg gctctttgcc
4741 acaactatct ctattggcta tatgccaata ctctgtcctt cagagactga cacggactct
4801 gtatttttac aggatggggt cccatttatt atttacaaat tcacatatac aacaacgccg
4861 tccccgtgc ccgcagtttt tattaaacat agcgtgggat ctccacgcga atctcgggta
4921 cgtgttccgg acatgggctc ttctccggta gcggcggagc ttccacatcc gagccctggt
4981 cccatgcctc cagcggctca tggtcgctcg gcagctcctt gctcctaaca gtggaggcca
5041 gacttaggca cagcacaatg cccaccacca ccagtgtgcc gcacaaggcc gtggcggtag
5101 ggtatgtgtc tgaaaatgag cgtggagatt gggctcgcac ggctgacgca gatggaagac
5161 ttaaggcagc ggcagaagaa gatgcaggca gctgagttgt tgtattctga taagagtcag
5221 aggtaactcc cgttgcggtg ctgttaacgg tggagggcag tgtagtcttga gcagtactcg
5281 ttgctgccgc gcgccaccc agacataata gctgacagac taacagactg ttccttttcca
5341 tgggtctttt ctgcagtcac cgtcggatca atcattratc tcgtgacttc ttcgtgtgtg
5401 gtgtttacct atatatctaa atttaatatt tcgttcatta aaatttaata tatttcgacg
5461 atgaatttct caaggatatt tttcttcgtg ttcttggctt gtcaacagtt
5521 tcggctgcgc cagagccgaa aggtacccag gtgcagctgc aggagtcggg gggaggcttg
5581 gtaaagccgg gggggtccct tagagtctcc tgtgcagcct ctggattcac ttccagaaac
5641 gcctggatga gctgggtccg ccaggctcca gggaaggggc tggagtgggt cggccgtatt
5701 aaaagcaaaa ttgatggtgg gacaacagac tatgctgcac ccgtgaaagg cagattcacc
5761 atctcaagag atgattcaaa aaacacgtta tatctgcaaa tgaatagcct gaaagccgag
5821 gacacagccg tatattactg taccacgggg attatgataa catttgtggg agttatccct
5881 ccccgaatt ggggccaggg aaccctggtc acagtctcct cagcctccac caagggccca
5941 tcggtcttcc ccctggcacc ctcctccaag agcacctctg ggggcacagc ggccctgggc
6001 tgcctggtca aggactactt cccgaaccg gtgacggtgt cgtggaactc aggcgccctg
6061 accagcggcg tgcacacctt tccggctgtc ctacagtcct caggactcta cttccttagc
6121 aacgtggtga ccgtgccctc cagcagcttg ggcacccaga cctacatctg caacgtgaat
6181 cacaagccca gcaacaccaa ggtggacaag aaagttgagc ccaaatcttg tgacaaaact
6241 cacacatgcc caccgtgccc agcacctgaa ctcctggggg gaccgtcagt cttcctcttc
6301 ccccaaaaac ccaaggacac cctcatgatc tcccggaccc ctgaggtcac atgcgtggtg
6361 gtggacgtga gccacgaaga ccctgaggtc aagttcaact ggtacgtgga cggcgtggag
6421 gtgcataatg ccaagacaaa gccgcgggag gagcagtaca acagcacgta ccgtgtggtc
6481 agcgtcctca ccgtcctgca ccaggactgg ctgaatggca aggagtacaa gtgcaaggtc
6541 tccaacaaag ccctcccagc ccccatcgag aaaaccatct ccaaagccaa agggcagccc
6601 cgagaaccac aggtgtacac cctgcccca tcccgggatg agctgaccaa gaaccaggtc
6661 agcctgacct gcctggtcaa aggcttctat cccagcgaca tcgccgtgga gtgggagagc
6721 aatgggcagc cggagaacaa ctacaagacc acgcctcccg tgctggactc cgacggctcc
6781 ttcttcctct acagcaagct caccgtggac aagagcaggt ggcagcaggg gaacgtcttc
6841 tcatgctccg tgatgcatga ggctctgcac aaccactaca cgcagaagag cctctccctg

```
6901 tctccgggta aagcgccaga gccgaagctt tcctatgagc tgacacagcc accctcggtg
6961 tcagtgtccc caggacaaac ggccaggatc acctgctctg gagatgcatt gccagaaaaa
7021 tatgtttatt ggtaccagca gaagtcaggc caggcccctg tggtggtcat ctatgaggac
7081 agcaaacgac cctccgggat ctctctggct ccagctcagg gacaatggcc
7141 accttgacta tcagtggggc ccaggtggaa gatgaaggtg actactactg ttactcaact
7201 gacagcagtg gttatcatag ggaggtgttc agcggaggga ccaagctgac cgtcctaggt
7261 cagcccaagg ctgccccctc ggtcactctg ttcccaccct cctctgagga gctttcaagcc
7321 aacaaggcca cactggtgtg tctcataagt gactcctacc cgggagccgt gacagtggcc
7381 tggaaggcag atagcagccc cgtcaaggcg ggagtggaga ccaccacacc ctccaaacaa
7441 agcaacaaca agtacgcggc cagcagctac ctgagcctga cgcttgagca gtggaagtcc
7501 cacaaaagct acagctgcca ggtcacgcat gaagggagca ccgtggagaa gacagtggcc
7561 cctgcagaat gttcaccgcg gagggaggga agggcccttt ttgaaggggg aggaaacttc
7621 gcgccatgac tcctctcgtg ccccccgcac ggaacactga tgtgcagagg gccctctgcc
7681 attgctgctt cctctgccct tcctgtcaac tctgaatgtg gcttctttgc tactgccaca
7741 gcaagaaata aaatctcaac atctaaatgg gtttcctgag attttcaag agtcgttaag
7801 cacattcctt ccccagcacc cctgctgca ggccagtgcc aggcaccaac ttggctactg
7861 ctgcccatga gagaaatcca gttcaatatt ttccaaagca aaatggatta catatgccct
7921 agatcctgat taacaggtgt ttgtgattat ctgtgctttc gcttcaccca cattatccca
7981 ttgcctcccc tcgaggggag gcccggtacc caattcgccc tatagtgagt cgtattacgc
8041 gcgctcactg gccgtcgttt tacaacgtcg tgactgggaa aaccctggcg ttacccaact
8101 taatcgcctt gcagcacatc ccccttccgc cagctggcgt aatagcgaag aggcccgcac
8161 cgatcgccct tcccaacagt tgcgcagcct gaatggcgaa tggaaattgt aagcgttaat
8221 atttgttaa aattcgcgtt aaatttttgt taaatcagct cattttttaa ccaataggcc
8281 gaaatcggca aaatccctta taaatcaaaa gaatagaccg agatagggtt gagtgttgtt
8341 ccagtttgga acaagagtcc actattaaag aacgtggact ccaacgtcaa agggrgaaaa
8401 acgctctatc agggcgatgg cccactactc cgggatcata tgacaagatg tgtatccacc
8461 ttaacttaat gatttttacc aaaatcatta ggggattcat cagtgctcag ggtcaacgag
8521 aattaacatt ccgtcaggaa agcttatgat gatgatgtgc ttaaaaactt actcaatggc
8581 tggttatgca tatcgcaata catgcgaaaa acctaaaaga gcttgccgat aaaaaaggcc
8641 aatttattgc tatttaccgc ggcttttttat tgagcttgaa agataaataa aatagatagg
8701 ttttatttga agctaaatct tctttatcgt aaaaaatgcc ctcttgggtt atcaagaggg
8761 tcattatatt tcgcggaata acatcatttg gtgacgaaat aactaagcac ttgtcctcctg
8821 tttactcccc tgagcttgag gggttaacat gaaggtcatc gatagcagga taataataca
8881 gtaaaacgct aaaccaataa tccaaatcca gccatcccaa attggtagtg aatgattata
8941 aataacagca aacagtaatg ggccaataac acggttgcta ttggtaaggc tcaccaataa
9001 tccctgtaaa gcaccttgct gatgactctt tgtttggata gacatcactc cctgtaatgc
9061 aggtaaagcg atcccaccac cagccaataa aattaaaaca gggaaaacta accaaccttc
9121 agatataaac gctaaaaagg caaatgcact actatctgca ataaatccga gcagtactgc
9181 cgttttttcg cccatttagt ggctattctt cctgccacaa aggcttggaa tactgagtgt
```

```
 9241 aaaagaccaa gacccgtaat gaaaagccaa ccatcatgct attcatcatc acgatttctg
 9301 taatagcacc acaccgtgct ggattggcta tcaatgcgct gaaataataa tcaacaaatg
 9361 gcatcgttaa ataagtgatg tataccgatc agctttgttt ccctttagtg agggttaatt
 9421 gcgcgcttgg cgtaatcatg gtcatagctg tttcctgtgt gaaattgtta tccgctcaca
 9481 attccacaca acatacgagc cggaagcata aagtgtaaag cctgggggtgc ctaatgagtg
 9541 agctaactca cattaattgc gttgcgctca ctgcccgctt tccagtcggg aaacctgtcg
 9601 tgccagctgc attaatgaat cggccaacgc gcggggagag gcggtttgcg tattgggcgc
 9661 tcttccgctt cctgctcac tgactcgctg cgctcggtcg ttcggctgcg gcgagcggta
 9721 tcagctcact caaaggcggt aatacggtta tccacagaat caggggataa cgcaggaaag
 9781 aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta aaaaggccgc gttgctggcg
 9841 tttttccata ggctccgccc ccctgacgag catcacaaaa atcgacgctc aagtcagagg
 9901 tggcgaaacc cgacaggact ataaagatac caggcgtttc cccctggaag ctccctcgtg
 9961 cgctctcctg ttcgacccct gccgcttacc ggatacctgt ccgcctttct cccttcggga
10021 agcgtggcgc tttctcatag ctcacgctgt aggtatctca gttcggtgta ggtcgttcgc
10081 tccaagctgg gctgtgtgca cgaaccccc gttcagcccg accgctgcgc cttatccggt
10141 aactatcgtc ttgagtccaa cccggtaaga cacgacttat cgccactggc agcagccact
10201 ggtaacagga ttagcagagc gaggtatgta ggcggtgcta cagagttctt gaagtggtgg
10261 cctaactacg gctacactag aaggacagta tttggtatct gcgctctgct gaagccagtt
10321 accttcggaa aaagagttgg tagctcttga tccggcaaac aaaccaccgc tggtagcggt
10381 ggtttttttg tttgcaagca gcagattacg cgcagaaaaa aaggatctca agaagatcct
10441 ttgatctttt ctacggggtc tgacgctcag tggaacgaaa actcacgtta agggattttg
10501 gtcatgagat tatcaaaaag gatcttcacc tagatccttt taaattaaaa atgaagtttt
10561 aaatcaatct aaagtatata tgagtaaact tggtctgaca gttaccaatg cttaatcagt
10621 gaggcaccta tctcagcgat ctgtctatct cgttcatcca tagttgcctg actccccgtc
10681 gtgtagataa ctacgatacg ggagggctta ccatctggcc ccagtgctgc aatgataccg
10741 cgagacccac gctcaccggc tccagattta tcagcaataa accagccagc cggaagggcc
10801 gagcgcagaa gtggtcctgc aactttatcc gcctccatcc agtctattaa ttgttgccgg
10861 gaagctagag taagtagttc gccagttaat agtttgcgca acgttgttgc cattgctaca
10921 ggcatcgtgg tgtcacgctc gtcgtttggt atggcttcat tcagctccgg ttcccaacga
10981 tcaaggcgag ttacatgatc ccccatgttg tgcaaaaaag cggttagctc cttcggtcct
11041 ccgatcgttg tcagaagtaa gttggccgca gtgttatcac tcatggttat ggcagcactg
11101 cataattctc ttactgtcat gccatccgta agatgctttt ctgtgactgg tgagtactca
11161 accaagtcat tctgagaata gtgtatgcgg cgaccgagtt gctcttgccc ggcgtcaata
11221 cgggataata ccgcgccaca tagcagaact ttaaaagtgc tcatcattgg aaaacgttct
11281 tcggggcgaa aactctcaag gatcttaccg ctgttgagat ccagttcgat gtaacccact
11341 cgtgcaccca actgatcttc agcatctttt actttcacca gcgtttctgg gtgagcaaaa
11401 acaggaaggc aaaatgccgc aaaaaaggga ataagggcga cacggaaatg ttgaatactc
11461 atactcttcc tttttcaata ttattgaagc atttatcagg gttattgtct catgagcgga
11521 tacatatttg aatgtattta gaaaaataaa caaatagggg ttccgcgcac atttccccga
```

11581 aaagtgccac

SEQ ID NO:101 (pTnMCS (CMV-prepro-HCPro-Lys-CPA))

```
    1 ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga
   61 ccgctacact tgccagcgcc ctagcgcccg ctcctttcgc tttcttccct tcctttctcg
  121 ccacgttcgc cggcatcaga ttggctattg gccattgcat acgttgtatc catatcataa
  181 tatgtacatt tatattggct catgtccaac attaccgcca tgttgacatt gattattgac
  241 tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg
  301 cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt
  361 gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca
  421 atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc
  481 aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta
  541 catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac
  601 catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg
  661 atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg
  721 ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt
  781 acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg
  841 ccatccacgc tgttttgacc tccatagaag acaccgggac cgatccagcc tccgcggccg
  901 ggaacggtgc attggaacgc ggattccccg tgccaagagt gacgtaagta ccgcctatag
  961 actctatagg cacacccctt tggctcttat gcatgctata ctgtttttgg cttggggcct
 1021 atacacccgc gcttccttat gctataggtg atggtatagc ttagcctata ggtgtgggtt
 1081 attgaccatt attgaccact cccctattgg tgacgatact ttccattact aatccataac
 1141 atggctcttt gccacaacta tctctattgg ctatatgcca atactctgtc cttcagagac
 1201 tgacacggac tctgtatttt tacaggatgg ggtcccattt attatttaca aattcacata
 1261 tacaacaacg ccgtcccccg tgcccgcagt ttttattaaa catagcgtgg gatctccacg
 1321 cgaatctcgg gtacgtgttc cggacatggg ctcttctccg gtagcggcgg agcttccaca
 1381 tccgagccct ggtcccatgc ctccagcggc tcatggtcgc tcggcagctc cttgctccta
 1441 acagtggagg ccagacttag gcacagcaca atgcccacca ccaccagtgt gccgcacaag
 1501 gccgtggcgg tagggtatgt gtctgaaaat gagcgtggag attgggctcg cacggctgac
 1561 gcagatggaa gacttaaggc agcggcagaa gaagatgcag gcagctgagt tgttgtattc
 1621 tgataagagt cagaggtaac tcccgttgcg gtgctgttaa cggtggaggg cagtgtagtc
 1681 tgagcagtac tcgttgctgc cgcgcgcgcc accagacata atagctgaca gactaacaga
 1741 ctgttccttt ccatgggtct tttctgcagt caccgtcgga ccatgtgcga actcgatatt
 1801 ttacacgact ctctttacca attctgcccc gaattacact taaaacgact caacagctta
 1861 acgttggctt gccacgcatt acttgactgt aaaactctca ctcttaccga acttggccgt
 1921 aacctgccaa ccaaagcgag aacaaaacat aacatcaaac gaatcgaccg attgttaggt
 1981 aatcgtcacc tccacaaaga gcgactcgct gtataccgtt ggcatgctag ctttatctgt
 2041 tcgggcaata cgatgcccat tgtacttgtt gactggtctg atattcgtga gcaaaaacga
 2101 cttatggtat tgcgagcttc agtcgcacta cacggtcgtt ctgttactct ttatgagaaa
```

2161 gcgtcccgc ttcagagca atgttcaaag aaagctcatg accaatttct agccgacctt
2221 gcgagcattc taccgagtaa caccacaccg ctcattgtca gtgatgctgg ctttaaagtg
2281 ccatggtata aatccgttga gaagctgggt tggtactggt taagtcgagt aagaggaaaa
2341 gtacaatatg cagacctagg agcggaaaac tggaaaccta tcagcaactt acatgatatg
2401 tcatctagtc actcaaagac ttaggctat aagaggctga ctaaaagcaa tccaatctca
2461 tgccaaattc tattgtataa atctcgctct aaaggccgaa aaaatcagcg ctcgacacgg
2521 actcattgtc accacccgtc acctaaaatc tactcagcgt cggcaaagga gccatgggtt
2581 ctagcaacta acttacctgt tgaaattcga acacccaaac aacttgttaa tatcttatcg
2641 aagcgaatgc agattgaaga aaccttccga gacttgaaaa gtcctgccta cggactaggc
2701 ctacgccata gccgaacgag cagctcagag cgttttgata tcatgctgct aatcgccctg
2761 atgcttcaac taacatgttg gcttgcgggc gttcatgctc agaaacaagg ttgggacaag
2821 cacttccagg ctaacacagt cagaaatcga aacgtactct caacagttcg cttaggcatg
2881 gaagttttgc ggcattctgg ctacacaata acaagggaag acttactcgt ggctgcaacc
2941 ctactagctc aaaatttatt cacacatggt tacgcttgg ggaaattatg agcggatcgc
3001 tctagagcga tccgggatct cgggaaaagc gttgggacc aaaggtgcct tttatcatca
3061 ctttaaaaat aaaaaacaat tactccagtgc ctgttataag cagcaattaa ttatgattga
3121 tgcctacatc acaacaaaaa ctgatttaac aaatggttgg tctgccttag aaagtatatt
3181 tgaacattat cttgattata ttattgataa taataaaaac cttatcccta tccaagaagt
3241 gatgcctatc attggttgga atgaacttga aaaaaattag ccttgaatac attactggta
3301 aggtaaacgc cattgtcagc aaattgatcc aagagaacca acttaaagct ttcctgacgg
3361 aatgttaatt ctcgttgacc ctgagcactg atgaatcccc taatgatttt ggtaaaaatc
3421 attaagttaa ggtggataca catcttgtca tatgatcccg gtaatgtgag ttagctcact
3481 cattaggcac cccaggcttt acactttatg cttccggctc gtatgttgtg tggaattgtg
3541 agcggataac aatttcacac aggaaacagc tatgaccatg attacgccaa gcgcgcaatt
3601 aaccctcact aaagggaaca aaagctggag ctccaccgcg gtggcggccg ctctagaact
3661 agtggatccc ccgggctgca ggaattcgat atcaagctta tcgataccgc tgacctcgag
3721 catcagattg gctattggcc attgcatacg ttgtatccat atcataatat gtacatttat
3781 attggctcat gtccaacatt accgccatgt tgacattgat tattgactag ttattaatag
3841 taatcaatta cggggtcatt agttcatagc ccatatatgg agttccgcgt tacataactt
3901 acggtaaatg gcccgcctgg ctgaccgccc aacgacccgc gcccattgac gtcaataatg
3961 acgtatgttc ccatagtaac gccaatagggg actttccatt gacgtcaatg ggtggagtat
4021 ttacggtaaa ctgcccactt ggcagtacat caagtgtatc atatgtcaag tacgccccct
4081 attgacgtca atgacggtaa atggcccgcc tggcattatg cccagtacat gaccttatgg
4141 gactttccta cttggcagta catctacgta ttagtcatcg ctattaccat ggtgatgcgg
4201 ttttggcagt acatcaatgg gcgtggatag cggtttgact cacggggatt tccaagtctt
4261 caccccattg acgtcaatgg gagtttgttt tggcaccaaa atcaacgggg ctttccaaaa
4321 tgtcgtaaca actccgcccc attgacgcaa atgggcggta ggcgtgtacg gtgggaggtc
4381 tatataagca gagctcgttt agtgaaccgt cagatcgcct ggagacgcca tccacgctgt
4441 tttgacctcc atagaagaca ccgggaccga tccagcctcc gcggccgggg acggtgcatt

4501 ggaacgcgga ttcccgtgc caagagtgac gtaagtaccg cctatagact ctataggcac
4561 acccctttgg ctcttatgca tgctatactg ttttggcttt ggggcctata cacccccgct
4621 tccttatgct ataggtgatg gtatagctta gcctataggt gtgggttatt gaccattatt
4681 gaccactccc ctattggtga cgatactttc cattactaat ccataacatg gctctttgcc
4741 acaactatct ctattggcta tatgccaata ctctgtcctt cagagactga cacggactct
4801 gtatttttac aggatggggt cccatttatt atttacaaat tcacatatac aacaacgccg
4861 tcccccgtgc ccgcagtttt tattaaacat agcgtgggat ctccacgcga atctcgggta
4921 cgtgttccgg acatgggctc ttctcgggta gcggcggagc ttccacatcc gagccctggt
4981 cccatgcctc cagcggctca tggtcgctcg gcagctcctt gctcctaaca gtggaggcca
5041 gacttaggca cagcacaatg cccaccacca ccagtgtgcc gcacaaggcc gtggcggtag
5101 ggtatgtgtc tgaaaatgag cgtggagatt gggctcgcac ggctgacgca gatggaagac
5161 ttaaggcagc ggcagaagaa gatgcaggca gctgagttgt tgtatttctga taagagtcag
5221 aggtaactcc cgttgcggtg ctgttaacgg tggagggcag tgtagtcttga gcagtactcg
5281 ttgctgccgc gcgcgccacc agacataata gctgacagac taacagactg ttcctttcca
5341 tgggtctttt ctgcagtcac cgtcggatca atcatccatc tcgtgacttc ttcgtgtgtg
5401 gtgtttacct atatatctaa atttaatatt tcgtttatta aaatttaata tatttcgacg
5461 atgaatttct caaggatatt tttcttcgtg ttcgctttgg ttctggcttt gtcaacagtt
5521 tcggctgcgc cagagccgaa aggtacccag gtgcagctgc aggagtcggg gggaggcttg
5581 gtaaagccgg gggggtccct tagagtctcc tgtgcagcct ctggattcac tttcagaaac
5641 gcctggatga gctgggtccg ccaggctcca gggaaggggc tggagtgggt cggccgtatt
5701 aaaagcaaaa ttgatggtgg gacaacagac tatgctgcac ccgtgaaagg cagattcacc
5761 atctcaagag atgattcaaa aaacacgtta tatctgcaaa tgaatagcct gaaagccgag
5821 gacacagccg tatattactg taccacgggg attatgataa catttgggag agttatccct
5881 ccccgaatt ggggccaggg aaccctggtc accgtctcct cagcctccac caagggccca
5941 tcggtcttcc ccctggcacc ctcctccaag agcacctctg ggggcacagc ggccctgggc
6001 tgcctggtca aggactactt ccccgaaccg gtgacggtgt cgtggaactc aggcgccctg
6061 accagcggcg tgcacacctt tccggctgtc ctacagtcct caggactcta ctccttagc
6121 aacgtggtga ccgtgccctc cagcagcttg ggcacccaga cctacatctg caacgtgaat
6181 cacaagccca gcaaccaaa ggtggacaag aaagttgagc ccaaatcttg tgacaaaact
6241 cacacatgcc caccgtgccc agcacctgaa ctcctggggg gaccgtcagt cttcctcttc
6301 ccccaaaac ccaaggacac cctcatgatc tcccggaccc ctgaggtcac atgcgtggtg
6361 gtggacgtga gccacgaaga ccctgaggtc aagttcaact ggtacgtgga cggcgtggag
6421 gtgcataatg ccaagacaaa gccgcgggag gagcagtaca acagcacgta ccgtgtggtc
6481 agcgtcctca ccgtcctgca ccaggactgg ctgaatggca aggagtacaa gtgcaaggtc
6541 tccaacaaag ccctcccagc ccccatcgag aaaaccatct ccaaagccaa agggcagccc
6601 cgagaaccac aggtgtacac cctgcccccca tcccgggatg agctgaccaa gaaccaggtc
6661 agcctgacct gcctggtcaa aggcttctat cccagcgaca tcgccgtgga gtgggagagc
6721 aatgggcagc cggagaacaa ctacaagacc acgcctcccg tgctggactc cgacggctcc
6781 ttcttcctct acagcaagct caccgtggac aagagcaggt ggcagcaggg gaacgtcttc

```
6841 tcatgctccg tgatgcatga ggctcctgcac aaccactaca cgcagaagag cctctccctg
6901 tctccgggta aagcgccaga gccgaaaaag ctttcctatg agctgacaca gccaccctcg
6961 gtgtcagtgt ccccaggaca aacggccagg atcacctgct ctggagatgc attgccagaa
7021 aaatatgttt attggtacca gcagaagtca ggccaggccc ctgtggtggt catctatgag
7081 gacagcaaac gaccctccgg gatccctgag agattctctg gctccagctc agggacaatg
7141 gccaccttga ctatcagtgg ggcccaggtg gaagatgaag gtgactacta ctgttactca
7201 actgacagca gtggttatca tagggaggtg ttcagcggag ggaccaagct gaccgtccta
7261 ggtcagccca aggctgcccc ctcggtcact ctgttcccac ctcctctga ggagcttcaa
7321 gccaacaagg ccacactggt gtgtctccata agtgactcct acccgggagc cgtgacagtg
7381 gcctggaagg cagatagcag cccgtcaag gcgggagtgg agaccaccac accctccaaa
7441 caaagcaaca acaagtacgc ggccagcagc tacctgagcc tgacgcttga gcagtggaag
7501 tcccacaaaa gctacagctg ccaggtcacg catgaaggga gcaccgtgga gaagacagtg
7561 gcccctgcag aatgttcacc gcggagggag ggaagggccc ttttgaagg gggaggaaac
7621 ttcgcgccat gactcctctc gtgcccccg cacgga-cac tgatgtgcag agggccctct
7681 gccattgctg cttcctctgc ccttcctcgt cactc:gaat gtggcttctt tgctactgcc
7741 acagcaagaa ataaaatctc aacatctaaa tgggtttcct gagatttttc aagagtcgtt
7801 aagcacattc ctccccagc acccctgct gcaggccagt gccaggcacc aacttggcta
7861 ctgctgccca tgagagaaat ccagttcaat attttccaaa gcaaaatgga ttacatatgc
7921 cctagatcct gattaacagg tgttttgtat tatctgtgct ccctatagtg agtcgtatta
7981 ccattgcctc ccctcgaggg ggggcccggt acccaattcg ccctatagtg agtcgtatta
8041 cgcgcgctca ctggccgtcg ttttacaacg tcgtgactgg gaaaccctg gcgttaccca
8101 acttaatcgc cttgcagcac atccccctttt cgccagctgg cgtaatagcg aagaggcccg
8161 caccgatcgc ccttcccaac agttgcgcag cctgaatggc gaatggaaat tgtaagcgtt
8221 aatattttgt taaaattcgc gttaaatttt tgttaaatca gctcatttt taaccatag
8281 gccgaaatcg gcaaaatccc ttataaatca aaagaataga ccgagatagg gttgagtgtt
8341 gttccagttt ggaacaagag tccactatta aagaacgtgg actccaacgt caaagggcga
8401 aaaaccgtct atcagggcga tggcccacta ctccgggatc atatgacaag atgtgtatcc
8461 accttaactt aatgatttct accaaaatca ttagggatt catcagtgct cagggtcaac
8521 gagaattaac attccgtcag gaaagcttat gatgatgatg tgcttaaaaa cttactcaat
8581 ggctggttat gcatatcgca atacatgcga aaaacctaaa agagcttgcc gataaaaaag
8641 gccaatttat tgctatttac cgcggctttt tattgagctt gaaagataaa taaaatagat
8701 aggttttatt tgaagctaaa tcttctttat cgtaaaaaat gccctcttgg gttatcaaga
8761 gggtcattat atttcgcgga ataacatcat ttggtgacga aataactaag cacttgtctc
8821 ctgtttactc ccctgagctt gaggggttaa catgaaggtc atcgatagca ggataataat
8881 acagtaaaac gctaaaccaa taatccaaat ccagccatcc caaattggta gtgaatgatt
8941 ataaataaca gcaaacagta atgggccaat aacaccggtt gcattggtaa ggctcaccaa
9001 taatcctgt aaagcacctt gctgatgact ctttgtttgg atagacatca ctccctgtaa
9061 tgcaggtaaa gcgatcccac caccagccaa taaaattaaa acagggaaaa ctaaccaacc
9121 ttcagatata aacgctaaaa aggcaaatgc actactatct gcaataaatc cgagcagtac
```

```
 9181 tgccgttttt tcgcccattt agtggctatt cttcctgcca caaaggcttg gaatactgag
 9241 tgtaaaagac caagacccgt aatgaaaagc caaccatcat gctattcatc atcacgattt
 9301 ctgtaatagc accacaccgt gctggattgg ctatcaatgc gctgaaataa taatcaacaa
 9361 atggcatcgt taaataagtg atgtataccg atcagctttt gttcccttta gtgagggtta
 9421 attgcgcgct tggcgtaatc atggtcataq ctgtttcctg tgtgaaattg ttatccgctc
 9481 acaattccac acaacatacg agccggaagc ataaagtgta aagcctgggg tgcctaatga
 9541 gtgagctaac tcacattaat tgcgttgcgc tcactgcccg ctttccagtc gggaaacctg
 9601 tcgtgccagc tgcattaatg aatcggccaa cgcgcgggga gaggcggttt gcgtattggg
 9661 cgctcttccg cttcctcgct cactgactcg ctgcgctcgg tcgttcggct gcggcgagcg
 9721 gtatcagctc actcaaaggc ggtaatacgg ttatccacag aatcagggga taacgcagga
 9781 aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc gtaaaaaggc cgcgttgctg
 9841 gcgtttttcc ataggctccg cccccctgac gagcatcaca aaaatcgacg ctcaagtcag
 9901 aggtggcgaa acccgacagg actataaaga taccaggcgt ttcccccctgg aagctccctc
 9961 gtgcgctctc ctgttccgac cctgccgctt accggatacc tgtccgcctt tctcccttcg
10021 ggaagcgtgg cgctttctca tagctcacgc tgtaggtatc tcagttcggt gtaggtcgtt
10081 cgctccaagc tgggctgtgt gcacgaaccc cccgttcagc ccgaccgctg cgccttatcc
10141 ggtaactatc gtcttgagtc caacccggta agacacgact tatcgccact ggcagcagcc
10201 actggtaaca ggattagcag agcgaggtat gtaggcggtg ctacagagtt cttgaagtgg
10261 tggcctaact acggctacac tagaaggaca gtatttggta tctgcgctct gctgaagcca
10321 gttaccttcg gaaaaagagt tggtagctct tgatccggca aacaaaccac cgctggtagc
10381 ggtggtttt ttgtttgcaa gcagcagatt acgcgcagaa aaaaaggatc tcaagaagat
10441 ccttgatct ttctacggg gtctgacgct cagtggaacg aaaactcacg ttaagggatt
10501 ttggtcatga gattatcaaa aaggatcttc acctagatcc ttttaaatta aaaatgaagt
10561 tttaaatcaa tctaaagtat atatgagtaa acttggtctg acagttacca atgcttnatc
10621 agtgaggcac ctatcttcagc gatctgtcta tttcgttcat ccatagttgc gcccagtgc tgcaatgata
10681 gtcgtgtaga taactacgat acgggagggc ttaccatctg gccccagtgc tgcaatgata
10741 ccgcgagacc cacgctcacc ggctccagat ttatcagcaa taaaccagcc agccggaagg
10801 gccgagcgca gaagtggtcc tgcaacttta tccgcctcca tccagtctat taattgttgc
10861 cgggaagcta gagtaagtag ttcgccagtt aatagtttgc gcaacgttgt tgccattgct
10921 acaggcatcg tggtgtcacg ctcgtcgttt ggtatggctt cattcagctc cggttcccaa
10981 cgatcaaggc gagttacatg atcccccatg ttgtgcaaaa aagcggttag ctccttcggt
11041 cctccgatcg ttgtcagaag taagttggcc gcagtgttat cactcatggt tatggcagca
11101 ctgcataatt ctcttactgt catgccatcc gtaagatgct tttctgtgac tggtgagtac
11161 tcaaccaagt cattctgaga atagtgtatg cggcgaccga gttgctcttg cccggcgtca
11221 atacgggata ataccgcgcc acatagcaga actttaaaag tgctcatcat tggaaaacgt
11281 tcttcggggc gaaaactctc aaggatctta ccgctgttga gatccagttc gatgtaaccc
11341 actcgtgcac ccaactgatc ttcagcatct tttactttca ccagcgtttc tgggtgagca
11401 aaaacaggaa ggcaaaatgc cgcaaaaaag ggaataaggg cgacacggaa atgttgaata
11461 ctcatactct tcctttttca atattattga agcatttatc agggttattg tctcatgagc
```

```
11521 ggatacatat ttgaatgtat ttagaaaaat aaacaaatag gggttccgcg cacatttccc
11581 cgaaaagtgc cac

SEQ ID NO:102 (pTnMod (CHOvep-prepro-HCPro-CPA))
    1 ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga
   61 ccgctacact tgccagcgcc ctagcgcccg ctcctttcgc tttcttccct tcctttctcg
  121 ccacgttcgc cggcatcaga ttggctattg gccattgcat acgttgtatc catatcataa
  181 tatgtacatt tatattggct catgtccaac attaccgcca tgttgacatt gattattgac
  241 tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg
  301 cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc ccgcccattt
  361 gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca
  421 atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc
  481 aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta
  541 catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac
  601 catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg
  661 atttccaagt ctcccacccca ttgacgtcaa tgggagtttg ctttggcacc aaaatcaacg
  721 ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt
  781 acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg
  841 ccatccacgc tgttttgacc tccatagaag acaccgggac cgatccagcc tccgcggccg
  901 ggaacggtgc attggaacgc ggattcccg tgccaagagt gacgtaagta ccgcctatag
  961 actctatagg cacacccctt tggctcttat gcatgctata ctgttttgg cttggggcct
 1021 atacacccgc gcttccttat gctataggtg atggtatagc ttagcctata ggtgtgggtt
 1081 attgaccatt attgaccact ccctattggg tgacgatact ttccattact aatccataac
 1141 atggctcttt gccacaacta tctctattgg ctatatgcca atactctgtc cttcagagac
 1201 tgacacggac tctgtatttt tacaggatgg ggtcccattt attatttaca aattcacata
 1261 tacaacaacg ccgtcccccg tgcccgcagt ttttattaaa catagcgtgg gatctccacg
 1321 cgaatctcgg gtacgttgttc cggacatggg ctcttctccg gtagcggcgg agcttccaca
 1381 tccgagcctt ggtcccatgc ctccagcggc tcatggtcgc tcggcagctc cttgctccta
 1441 acagtggagg ccagacttag gcacagcaca atgcccacca ccaccagtgt gccgcacaag
 1501 gccgtggcgg taggtatgt gtctgaaaat gagcgtggag attgggctcg cacggctgac
 1561 gcagatggaa gacttaaggc agcggcagaa gaagatgcag gcagctgagt tgttgtattc
 1621 tgataagagt cagaggtaac tcccgttgcg gtgctgttaa cggtggaggg cagtgtagtc
 1681 tgagcagtac tcgttgctgc cgcgcgccc accagacata atagctgaca gactaacaga
 1741 ctgttccttt ccatgggtct tttctgcagt caccgtcgga ccatgtgtga acttgatatt
 1801 ttacatgatt ctctttacca attctgcccc gaattacact taaaacgact caacagctta
 1861 acgttggctt gccacgcatt acttgactgt aaaactctca ctctttaccga acttggccgt
 1921 aacctgccaa ccaaagcgag aacaaaacat aacatcaaac gaatcgaccg attgttaggt
 1981 aatcgtcacc tccacaaaga gcgactcgct ggcatgctag ctttatctgt
 2041 tcgggcaata cgatgcccat tgtacttgtt gactggtctg atattcgtga gcaaaaacga
```

```
2101 cttatggtat tgcgagcttc agtcgcacta cacggtcgtt ctgttactct ttatgagaaa
2161 gcgttcccgc tttcagagca atgttcaaag aaagctcatg accaatttct agccgacctt
2221 gcgagcattc taccgagtaa caccacaccg ctcattgtca gtgatgctgg ctttaaagtg
2281 ccatggtata aatccgttga gaagctgggt tggtactggt taagtcgagt aagaggaaaa
2341 gtacaatatg cagacctagg agcggaaaac tggaaaccta tcagcaactt acatgatatg
2401 tcatctagtc actcaaagac tttaggctat aagaggctga ctaaaagcaa tccaatctca
2461 tgccaaattc tattgtataa atctcgctct aaaggccgaa aaaatcagcg ctcgacacgg
2521 actcattgtc accacccgtc acctaaaatc tactcagcgt cggcaaagga gccatgggtt
2581 ctagcaacta acttacctgt tgaaattcga acacccaaac aacttgttaa tatctattcg
2641 aagcgaatgc agattgaaga aaccttccga gacttgaaaa gtcctgccta cggactaggc
2701 ctacgccata gccgaacgag cagctcagag cgttttgata tcatgctgct aatcgccctg
2761 atgcttcaac taacatgttg gcttgcgggc gttcatgctc agaaacaagg ttgggacaag
2821 cacttccagg ctaacacagt cagaaatcga aacgtactct caacagttcg cttaggcatg
2881 gaagttttgc ggcattctgg ctacacaata acaaggtaag acttactcgt ggctgcaacc
2941 ctactagctc aaaatttatt cacacatggt tacgctttgg ggaaattatg ataatgatcc
3001 agatcacttc tggctaataa aagatcagag ctctagagat ctgtgtgttg gttttttgtg
3061 gatctgctgt gccttctagt tgccagccat ctgttgtttg cccctccccc gtgccttcct
3121 tgaccctgga aggtgccact cccactgtcc tttcctaata aaatgaggaa attgcatcgc
3181 attgtctgag taggtgtcat tctattctgg ggggtggggt ggggcagcac agcaaggggg
3241 aggattggga agacaatagc aggcatgctg gggatgcggt gggctctatg ggtacctctc
3301 tctctctctc tctctctctc tctctcggta cctctctctc tctctctctc
3361 tctctctctc tctctctctc tcggtaccag gtgctgaaga attgacccgg tgaccaaagg
3421 tgcttttat catcacttta aaaataaaaa acaattactc agtgcctgtt ataagcagca
3481 attaattatg attgatgcct acatcacaac aaaaactgat ttaacaaatg gttggtctgc
3541 cttagaaagt atatttgaac attatcttga ttatattatt gataataata aaaaccttat
3601 ccctatccaa gaagtgatgc ctatcattgg ttggaatgaa cttgaaaaaa attagccttg
3661 aatacattac tggtaaggta aacgccattg tcagcaaatt gatccaagag aaccaactta
3721 aagctttcct gacggaatgt taattctcgt tgaccctgag cactgatgaa tccccttaatg
3781 atttggtaa aaatcattaa gttaaggtgg atacacatct tgtcatatga tcccggtaat
3841 gtgagttagc tcactcatta ggcaccccag gctttacact ttatgcttcc ggctcgtatg
3901 ttgtgtggaa ttgtgagcgg ataacaattt cacacaggaa acagctatga ccatgattac
3961 gccaagcgcg caattaaccc tcactaaagg gaacaaaagc tggagctcca ccgcggtggc
4021 ggccgctcta gaactagtgg atcccccggg ctgcaggaat tcgatatcaa gcttatcgat
4081 accgctgacc tcgagctgca gaaaaatgcc aggtggacta tgaactcaca tccaaaggag
4141 cttgacctga tacctgattt tcttcaaaca ggggaaacaa cacaatccca caaaacagct
4201 cagagagaaa ccatcactga tggctacagc accaaggtat gcaatggcaa tccattcgac
4261 attcatctgt gacctgagca aaatgatttc tctctccatg aatggttgct tctttccctc
4321 atgaaaaggc aatttccaca ctcacaatat gcgacaaaga caaacagaga acaattaatg
4381 tgctccttcc taatgtcaaa attgtagtgg caaagaggag aacaaaatct caagtttcga
```

```
4441 gtaggtttta gtgattggat aagaggcttt gacctgtgag ctcacctgga cttcatatcc
4501 tttggataa aaagtgcttt tataactttc aggtctccga gtctttattc atgagactgt
4561 tggtttaggg acagacccac aatgaaatgc ctggcatagg aaagggcagc agagccttag
4621 ctgacctttt cttgggacaa gcattgtcaa acaatgtgtg acaaaactat ttgtactgct
4681 ttgcacagct gtgctgggca gggcgatcca ttgccaccta tcccaggtaa ccttccaact
4741 gcaagaagat tgttgcttac tctctctaga aagcttctgc agactgacat gcatttcata
4801 ggtagagata acatttactg ggaagcacat ctatcatcac aaaaagcagg caagatttc
4861 agacttcctt agtggctgaa atagaagcaa aagacgtaat taaaaacaaa atgaaacaaa
4921 aaaaatcagt tgatacctgt ggtgtagaca tccagcaaaa aaatattatt tgcactacca
4981 tcttgtctta agtcctcaga cttagcaagg agaatgtaga tttccacagt atatatgttt
5041 tcacaaaagg aaggagagaa acaaaagaaa atggcactga ctaaacttca gctagtggta
5101 taggaaagta atctgcttа acagagattg cagtgatctc tatgtatgtc ctgaagaatt
5161 atgttgtact ttttcccccc attttaaat caaacagtgc tttacagagg tcagaatggt
5221 ttcttttactg tttgtcaatt ctattatttc aatacaaaac aatagcttct ataactgaaa
5281 tatatttgct attgtatatt atgattgtcc ctcgaaccat gaacactcct ccagctgaat
5341 ttcacaattc ctctgtcatc tgccaggcca ttaagtttatt catggaagat ctttgaggaa
5401 cactgcaagt tcatatcata aacacatttg aaattgagta ttggtttgca ttgtatggag
5461 ctatgttttg ctgtatcctc agaaaaaaag tttgttataa agcattccaca cccataaaaa
5521 gatagattta aatattccag ctataggaaa gaaagtgcgt ctgctcttca ctctagtctc
5581 agttggctcc ttcacatgca tgcttctttа ttctcctat tttgtcaaga aaataatagg
5641 tcacgtcttg ttctcactta tgtcctgcct agcatggctc agatgcacgt tgtagataca
5701 agaaggatca aatgaaacag acttctggtc tgttacctac aaccatagta ataagcacac
5761 taactaataa ttgctaatta tgtttttccat ctctaaggtt cccatatttt tctgtttttct
5821 taaagatccc attatctggt tgtaactgaa gctcaatgga acatgagcaa tatttcucag
5881 tcttctctcc catccaacag tcctgatgga ttagcagaac aggcagaaaa cacattgtta
5941 cccagaatta aaaactaata tttgctctcc attcaatcca aaatggacct attgaaacta
6001 aaatctaacc caatcccatt aaatgatttc tatggtgtca aaggtcaaac ttctgaaggg
6061 aacctgtggg tgggtcacaa ttcaggctat atattcccca gggctcagcc agtggatcaa
6121 tcattcatct cgtgacttct tcgtgtgtgg tgtttaccta tatatctaaa tttaatattt
6181 cgtttattaa aatttaatat atttcgacga tgaatttctc aaggatattt ttcttcgtgt
6241 tcgctttggt tctggctttg tcaacagttt cggctgcgcc agagccgaaa ggtacccagg
6301 tgcagctgca ggagtcgggg ggaggcttgg taaagccggg ggggtccctt agagtctcct
6361 gtgcagcctc tggattcact ttcagaaacg cctggatgag ctgggtccgc caggctccag
6421 ggaaggggct ggagtgggtc ggccgtatta aaagcaaaat tgatggtggg acaacagact
6481 atgctgcacc cgtgaaaggc agattcacca tctcaagaga tgattcaaaa aacacgttat
6541 atctgcaaat gaatagcctg aaagccgagg acacagccgt atattactgt accacgggga
6601 ttatgataac atttgggggа gttatcccctc ccccgaattg gggccaggga accctggtca
6661 ccgtctcctc agcctccacc aagggcccat cggtcttccc cctggcaccc tcctccaaga
6721 gcacctctgg gggcacagcg gccctgggct gcctggtcaa ggactacttc cccgaaccgg
```

```
6781 tgacggtgtc gtggaactca ggcgccctga ccagcggcgt gcacaccttt ccggctgtcc
6841 tacagtcctc aggactctac ttccttagca acgtggtgac cgtgccctcc agcagcttgg
6901 gcacccagac ctacatctgc aacgtgaatc acaagcccag caacaccaag gtggacaaga
6961 aagttgagcc caaatcttgt gacaaaactc acacatgccc accgtgccca gcacctgaac
7021 tcctgggggg accgtcagtc ttcctcttcc ccccaaaacc caaggacacc ctcatgatct
7081 cccggacccc tgaggtcaca tgcgtggtgg tggacgtgag ccacgaagac cctgaggtca
7141 agttcaactg gtacgtggac ggcgtggagg tgcataatgc caagacaaag ccgcgggagg
7201 agcagtacaa cagcacgtac cgtgtggtca gcgtcctcac cgtcctgcac caggactggc
7261 tgaatggcaa ggagtacaag tgcaaggtct ccaacaaagc cctcccagcc cccatcgaga
7321 aaaccatctc caaagccaaa gggcagcccc gagaaccaca ggtgtacacc ctgccccccat
7381 cccgggatga gctgaccaag aaccaggtca gcctgacctg cctggtcaaa ggcttctatc
7441 ccagcgacat cgccgtggag tgggagagca atgggcagcc ggagaacaac tacaagacca
7501 cgcctcccgt gctggactcc tcttcctcta cagcaagctc accgtggaca
7561 agagcaggtg gcagcagggg aacgtcttct catgctccgt gatgcatgag gctctgcaca
7621 accactacac gcagaagagc ctctccctgt ctccgggtaa agcgccagag ccgaagcttt
7681 cctatgagct gacacagcca ccctcgggtgt cagtgtcccc aggacaaacg gccaggatca
7741 cctgcttctgg agatgcattg ccagaaaaat atgtttattg gtaccagcag aagtcaggcc
7801 aggcccctgt ggtggtcatc tatgaggaca gcaaacgacc ctccgggatc cctgagagat
7861 tctctggctc cagctcaggg acaatggcca ccttgactat cagtgggggcc caggtggaag
7921 atgaaggtga ctactactgt tactcaactg acagcagtgg ttatcatagg gaggtgttca
7981 gcggagggac caagctgacc gtcctaggtc agcccaaggc tgccccctcg gtcactctgt
8041 tcccacccctc ctctgaggag cttcaagcca acaaggccac actggtgtgt ctcataagtg
8101 actcctaccc gggagccgtg acagtggcct ggaaggcaga tagcagcccc gtcaaggcgg
8161 gagtggagac caccacaccc tccaaacaaa gcaacaacaa gtacgcggcc agcagctacc
8221 tgagcctgac gcttgagcag tggaagtccc acaaaagcta cagctgccag gtcaccatg
8281 aagggagcac cgtggagaag acagtggccc ctgcagaatg ttcaccgcgg agggagggaa
8341 gggccctttt tgaagggga ggaaacttcg cgccatgact cctctcgtgc cccccgcacg
8401 gaacactgat gtgcagaggg ccctctgcca ttgctgcttc ctctgccctt cctcgtcact
8461 ctgaatgtgg cttctttgct actgccacag caagaaataa aatctcaaca tctaaatggg
8521 ttcctgaga ttttcaaga gtcgttaagc acattccttc cccagcaccc cttgctgcag
8581 gccagtgcca ggcaccaact tggctactgc tgcccatgag agaaatccag ttcaatattt
8641 tccaaagcaa aatggattac atatgcccta gatcctgatt aacaggtgtt ttgtattatc
8701 tgtgctttcg cttcacccac attatcccat tgcctcccct cgaggggggg cccggtaccc
8761 aattcgccct atagtgagtc gtattacgcg cgtcactggc ccgtcgtttt acaacgtcgt
8821 gactgggaaa accctggcgt tacccaactt aatcgccttg cagcacatcc cctttcgcc
8881 agctggcgta atagcgaaga ggcccgcacc gatcgccctt cccaacagtt gcgcagcctg
8941 aatggcgaat ggaaattgta agcgttaata tttgttaaa attcgcgtta aattttgtt
9001 aaatcagctc attttttaac caataggccg aaatcggcaa aatcccttat aaatcaaaag
9061 aatagaccga gatagggttg agtgttgttc cagtttggaa caagagtcca ctattaaaga
```

```
9121 acgtggactc caacgtcaaa gggcgaaaaa ccgtctatca gggcgatggc ccactactcc
9181 gggatcatat gacaagatgt gtatccacct taacttaatg attttacca aaatcattag
9241 gggattcatc agtgctcagg gtcaacgaga attaacattc cgtcaggaaa gcttatgatg
9301 atgatgtgct taaaaactta ctcaatggct ggttatgcat atcgcaatac atgcgaaaaa
9361 cctaaaagag cttgccgata aaaaaggcca atttattgct atttaccgcg gcttttatt
9421 gagcttgaaa gataaataaa atagataggt tttatttgaa gctaaatctt ctttatcgta
9481 aaaaatgccc tcttgggtta tcaagagggt cattatattt cgcggaataa catcatttgg
9541 tgacgaaata actaagcact tgtctcctgt ttactccccc gagcttgagg ggttaacatg
9601 aaggtcatcg atagcaggat aataatacag taaaacgcta aaccaataat ccaaatccag
9661 ccatcccaaa ttggtagtga atgattataa ataacagcaa acagtaatgg gccaataaca
9721 ccggttgcat tggtaaggct caccaataat ccctgtaaag caccttgctg atgactcttt
9781 gtttggatag acatcactcc ctgtaatgca ggtaaagcga tcccaccacc agccaataaa
9841 attaaaacag ggaaaactaa ccaaccttca gatataaacg ctaaaaaggc aaatgcacta
9901 ctatctgcaa taaatccgag cagtactgcc gttttt-cgc ccatttagtg gctattcttc
9961 ctgccacaaa ggcttggaat actgagtgta aaagaccaag acccgtaatg aaaagccaac
10021 catcatgcta ttcatcatca cgatttctgt aatagcacca caccgtgctg gattggctat
10081 caatgcgctg aaataataat caacaaatgg catcgttaaa taagtgatgt ataccgatca
10141 gcttttgttc cctttagtga gggttaattg cgcgcttggc gtaatcatgg tcatagctgt
10201 ttcctgtgtg aaattgttat ccgctcacaa ttccacacaa catacgagcc ggaagcataa
10261 agtgtaaagc ctgggtgccc taatgagtga gctaactcac attaattgcg ttgcgctcac
10321 tgcccgcttt ccagtcggga aacctgtcgt gccagctgca ttaatgaatc ggccaacgcg
10381 cggggagagg cggtttgcgt attgggcgct cttccgcttc ctcgctcact gactcgctgc
10441 gctcggtcgt tcggctgcgg cgagcggtat cagctcactc aaaggcggta atacggttat
10501 ccacagaatc aggggataac gcaggaaaga acatgtgagc aaaaggccag caaaaggcca
10561 ggaaccgtaa aaaggccgcg ttgctggcgt ttttccataq gctccgcccc cctgacgagc
10621 atcacaaaaa tcgacgctca agtcagaggt ggcgaaaccc gacaggacta taaagatacc
10681 aggcgtttcc ccctggaagc tccctcgtgc gctctcctgt tccgaccctg ccgcttaccg
10741 gatacctgtc cgcctttctc ccttcgggaa gcgtggcgct ttctcatagc tcacgctgta
10801 ggtatctcag ttcggtgtag gtcgttcgct ccaagctggg ctgtgtgcac gaacccccg
10861 ttcagcccga ccgctgcgcc ttatccggta actatcgtct tgagtccaac ccggtaagac
10921 acgacttatc gccactggca gcagccactg gtaacaggat tagcagagcg aggtatgtag
10981 gcggtgctac agagttcttg aagtggtggc ctaactacgg ctacactaga aggacagtat
11041 ttggtatctg cgctctgctg aagccagtta ccttcggaaa aagagttggt agctcttgat
11101 ccggcaaaca aaccaccgct ggtagcggtg gtttttttgt ttgcaagcag cagattacgc
11161 gcagaaaaaa aggatctcaa gaagatcctt tgatcttttc tacggggtct gacgctcagt
11221 ggaacgaaaa ctcacgttaa gggattttgg tcatgagatt atcaaaaagg atcttcacct
11281 agatcctttt aaattaaaaa tgaagtttta aatcaatcta aagtatatat gagtaaactt
11341 ggtctgacag ttaccaatgc ttaatcagtg aggcacctat ctcagcgatc tgtctatttc
11401 gttcatccat agttgcctga ctccccgtcg tgtagataac tacgatacgg gagggcttac
```

```
11461 catctggccc cagtgctgca atgataccgc gagacccacg ctcaccggct ccagatttat
11521 cagcaataaa ccagccagcc ggaagggccg agcgcagaag tggtcctgca actttatccg
11581 cctccatcca gtctattaat tgttgccggg aagctagagt aagtagttcg ccagttaata
11641 gtttgcgcaa cgttgttgcc attgctacag gcatcgtggt gtcacgctcg tcgtttggta
11701 tggcttcatt cagctccggt tcccaacgat caaggcgagt tacatgatcc cccatgttgt
11761 gcaaaaaagc ggttagctcc ttcggtcctc cgatcgttgt cagaagtaag ttggccgcag
11821 tgttatcact catggttatg gcagcactgc ataattctct tactgtcatg ccatccgtaa
11881 gatgcttttc tgtgactggt gagtactcaa ccaagtcatt ctgagaatag tgtatgcggc
11941 gaccgagttg ctcttgcccg gcgtcaatac gggataatac cgcgccacat agcagaactt
12001 taaaagtgct catcattgga aaacgttctt cggggcgaaa actctcaagg atcttaccgc
12061 tgttgagatc cagttcgatg taacccactc gtgcacccaa ctgatcttca gcatctttta
12121 ctttcaccag cgtttctggg tgagcaaaaa caggaaggca aaatgccgca aaaaagggaa
12181 taagggcgac acggaaatgt tgaatactca tactcttcct ttttcaatat tattgaagca
12241 tttatcaggg ttattgtctc atgagcggat acatatttga atgtatttag aaaaataaac
12301 aaatagggggt tccgcgcaca tttccccgaa aagtgccac
```

SEQ ID NO:103 (pTnMod (CHOvep-prepro-HCPro-LYS-CPA))

```
    1 ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga
   61 ccgctacact tgccagcgcc ctagcgcccg ctcctttcgc tttcttccct tcctttctcg
  121 ccacgttcgc cggcatcaga ttggctattg gccattgcat acgttgtatc catatcataa
  181 tatgtacatt tatattggct catgtccaac attaccgcca tgttgacatt gattattgac
  241 tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg
  301 cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt
  361 gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca
  421 atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc
  481 aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta
  541 catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac
  601 catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg
  661 atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg
  721 ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt
  781 acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg
  841 ccatccacgc tgttttgacc tccatagaag acaccgggac cgatccagcc tccgcggccg
  901 ggaacggtgc attggaacgc ggattccccg tgccaagagt gacgtaagta ccgcctatag
  961 actctatagg cacacccctt tggctcttat gcatgctata ctgtttttgg cttggggcct
 1021 atacacccgc gcttccttat gctataggtg atggtatagc ttagcctata ggtgtgggtt
 1081 attgaccatt attgaccact cccctattgg tgacgatact ttccattact aatccataac
 1141 atggctcttt gccacaacta tctctattgg ctatatgcca atactctgtc cttcagagac
 1201 tgacacggac tctgtatttt tacaggatgg ggtcccattt attatttaca aattcacata
 1261 tacaacaacg ccgtcccccg tgcccgcagt ttttattaaa catagcgtgg gatctccacg
```

```
1321 cgaatctcgg gtacgtgttc cggacatggg ctcttctccg gtagcggcgg agcttccaca
1381 tccgagccct ggtcccatgc ctccagcggc tcatggtcgc tcggcagctc cttgctccta
1441 acagtggagg ccagacttag gcacagcaca atgcccacca ccaccagtgt gccgcacaag
1501 gccgtggcgg tagggtatgt gtcttgaaat gagcgtggag attgggctcg cacggctgac
1561 gcagatggaa gacttaaggc agcggcagaa gaagatgcag gcagctgagt tgttgtattc
1621 tgataagagt cagaggtaac tcccgttgcg gtgctgttaa cggtggaggg cagtgtagtc
1681 tgagcagtac tcgttgctgc cgcgcgcgcc accagacata atagctgaca gactaacaga
1741 ctgttccttt ccatgggtct ttcctgcagt caccgtcgga ccatgtgtga acttgatatt
1801 ttacatgatt ctctttacca atcctgcccc gaattacact taaaacgact caacagctta
1861 acgttggctt gccacgcatt acttgactgt aaaactctca ctctttaccga acttggccgt
1921 aacctgccaa ccaaagcgag aacaaaaact aacatcaaac gaatcgaccg attgttaggt
1981 aatcgtcacc tccacaaaga gcgactcgct gtataccgtt ggcatgctag ctttatctgt
2041 tcgggcaata cgatgcccat tgtacttgtt gactggtctg atattcgtga gcaaaaacga
2101 cttatggtat tgcgagcttc agtcgcacta cacggtcgtt ctgttactct ttatgagaaa
2161 gcgttcccgc tttcagagca atgttcaaag aaagcccatg accaatttct agccgacctt
2221 gcgagcattc taccgagtaa caccacaccg ctcattgtca gtgatgctgg ctttaaagtg
2281 ccatggtata aatccgttga gaagctgggt tggtactggt taagtcgagt aagaggaaaa
2341 gtacaatatg cagacctagg agcggaaaac tggaaaccta tcagcaactt acatgatatg
2401 tcatctagtc actcaaagac tttaggctat aagaggctga ctaaaagcaa tccaatctca
2461 tgccaaattc tattgtataa atctcgctct aaaggccgaa aaaatcagcg ctcgacacgg
2521 actcattgtc accacccgtc acctaaaatc tactcagcgt cggcaaagga gccatgggtt
2581 ctagcaacta acttacctgt tgaaattcga acacccaaac aacttgttaa tatctattcg
2641 aagcgaatgc agattgaaga aaccttccga gacttgaaaa gtcctgccta cggactaggc
2701 ctacgccata gccgaacgag cagctcagag cgttttgata tcatgctgct aatcgcctg
2761 atgcttcaac taacatgttg gcttgcgggc gttcatgctc agaaacaagg ttgggacaag
2821 cacttccagg ctaacacagt cagaaatcga aacgtactct caacagttcg cttaggcatg
2881 gaagttttgc ggcattctgg ctacacaata acaagggaag acttactcgt ggctgcaacc
2941 ctactagctc aaaatttatt cacacatggt tacgctttgg ggaaattatg ataatgatcc
3001 agatcacttc tggctaataa aagatcagag ctctagagat ctgtgtgttg gttttttgtg
3061 gatctgctgt gccttctagt tgccagccat ctgttgtttg ccccttcccc gtgccttcct
3121 tgaccctgga aggtgccact cccactgtcc tttcctaata aaatgaggaa attgcatcgc
3181 attgtctgag taggtgtcat tctattctgg ggggtggggt ggggcagcac agcaaggggg
3241 aggattgggg agacaatagc aggcatgctg gggatgcggt gggctctatg ggtacctctc
3301 tctcctctct tctctctctc tctctctctc tctctcggta cctctctctc tctctctctc
3361 tctctctctc tctctctctc tcggtaccag gtgctgaaga attgacccgg tgaccaaagg
3421 tgcctttttat catcacttta aaaataaaaa acaattactc agtgcctgtt ataagcagca
3481 attaattatg attgatgcct acatcacaac aaaaactgat ttaacaaatg gttggtctgc
3541 cttagaaagt atattgaac attatcttga ttatattatt gataataata aaaaccttat
3601 ccctatccaa gaagtgatgc ctatcattgg ttggaatgaa cttgaaaaaa attagccttg
```

3661 aatacattac tggtaaggta aacgccattg tcagcaaatt gatccaagag aaccaactta
3721 aagctttcct gacggaatgt taattctcgt tgaccctgag cactgatgaa tcccctaatg
3781 attttggtaa aaatcattaa gttaaggtgg atacacatct tgtcatatga tcccggtaat
3841 gtgagttagc tcactcatta ggcacccccag gctttacact ttatgcttcc ggctcgtatg
3901 ttgtgtggaa ttgtgagcgg ataacaattt cacacaggaa acagctatga ccatgattac
3961 gccaagcgcg caattaaccc tcactaaagg gaacaaaagc tggagctcca ccgcggtggc
4021 ggccgctcta gaactagtgg atcccccggg ctgcaggaat tcgatatcaa gcttatcgat
4081 accgctgacc tcgagctgca gaaaaatgcc aggtggacta tgaactcaca tccaaaggag
4141 cttgacctga tacctgattt tcttcaaaca ggggaaacaa cacaatccca caaaacagct
4201 cagagagaaa ccatcactga tggctacagc accaaggtat gcaatggcaa tccattcgac
4261 attcatctgt gacctgagca aaatgatttc tctctccatg aatggttgct tctttccctc
4321 atgaaaaggc aatttccaca ctcacaatat gcgacaaaga caaacagaga acaattaatg
4381 tgctccttcc taatgtcaaa attgtagtgg caaagaggag aacaaaatct caagttcttga
4441 gtaggtttta gtgattggat aagaggcttt gacctg-gag ctcacctgga cttcatatcc
4501 tttggataa aaagtgcttt tataacttcc aggtctccga gtctttattc atgagactgt
4561 tggtttaggg acagacccac aatgaaatgc ctggcatagg aaagggcagc agagccttag
4621 ctgaccttt cttgggacaa gcattgtcaa acaatgtgtg acaaaactat ttgtactgct
4681 ttgcacagct gtgctgggca gggcgatcca ttgccaccta tcccaggtaa ccttccaact
4741 gcaagaagat tgttgcttac tctctctaga aagcttctgc agactgacat gcatttcata
4801 ggtagagata acatttactg ggaagcacat ctatcatcac aaaaagcagg caagatttcc
4861 agactttctt agtggctgaa atagaagcaa aagacgtaat taaaaacaaa atgaaacaaa
4921 aaaaatcagt tgatacctgt ggtgtagaca tccagcaaaa aaatattatt tgcactacca
4981 tcttgtctta agtcctcaga cttagcaagg agaatgtaga tttccacagt atatatgttt
5041 tcacaaaagg aaggagagaa acaaaagaaa atggcactga ctaaacttca gctagtggta
5101 taggaaagta attctgctta acagagattg cagtgatcttc tatgtatgtc ctgaagaatt
5161 atgttgtact ttttcccccc attttaaat caaacagtgc tttacagagg tcagaatggt
5221 ttctttactg ttgtcaatt ctattatttc aatacagaac aatagcttct ataactgaaa
5281 tatatttgct attgtatatt atgattgtcc ctcgaaccat gaacactcct ccagctgaat
5341 ttcacaattc ctctgtcatc tgccaggcca ttaagttatt catggaagat ctttgaggaa
5401 cactgcaagt tcatatcata aacacatttg aaatgagta ttggtttgca ttgtatggag
5461 ctatgttttg ctgtatcctc agaaaaaaag ttgttataa agcattcaca cccataaaaa
5521 gatagattta aatattccag ctataggaaa gaaagtgcgt ctgctcttca ctctagtctc
5581 agttggctcc ttcacatgca tgcttcttta ttctccctat tttgtcaaga aaataatagg
5641 tcacgtcttg ttctcactta tgtcctgcct agcatggctc agatgcacgt tgtagataca
5701 agaaggatca aatgaaacag acttctggtc tgttacctac aaccatagta ataagcacac
5761 taactaataa ttgctaatta tgttttccat ctctaaggtt cccatatttt tctgttttct
5821 taaagatccc attatctggt tgtaactgaa gctcaatgga acatgagcaa tatttcccag
5881 tcttctctcc catccaacag tcctgatgga ttagcagaac aggcagaaaa cacattgtta
5941 cccagaatta aaaactaata tttgctctcc attcaatcca aaatggacct attgaaacta

6001 aaatctaacc caatcccatt aaatgatttc tatggtgtca aaggtcaaac ttctgaaggg
6061 aacctgtggg tgggtcacaa ttcaggctat atattcccca gggctcagcc agtggatcaa
6121 tcattcatct cgtgacttct tcgtgtgtgg tgtttaccta tatatctaaa tttaatattt
6181 cgtttattaa aatttaatat atttcgacga tgaatttctc aaggatattt ttcttcgtgt
6241 tcgctttggt tctggctttg tcaacagttt cggctgcgcc agagccgaaa ggtacccagg
6301 tgcagctgca ggagtcgggg ggaggcttgg taaagccggg ggggtccctt agagtctcct
6361 gtgcagcctc tggattcact ttcagaaacg cctggatgag ctgggtccgc caggctccag
6421 ggaaggggct ggagtgggtc ggccgtatta aaagcaaaat tgatggtggg acaacagact
6481 atgctgcacc cgtgaaaggc agattcacca tctcaagaga tgattcaaaa aacacgttat
6541 atctgcaaat gaatagcctg aaagccgagg acacagccgt atattactgt accacgggga
6601 ttatgataac atttgggggа gttatccctc ccccgaattg gggccaggga accctggtca
6661 ccgtctcctc agcctccacc aagggcccat cggtcttccc cctggcaccc tcctccaaga
6721 gcacctctgg gggcacagcg gccctgggct gcctggtcaa ggactacttc cccgaaccgg
6781 tgacggtgtc gtggaactca ggcgccctga ccagcgtcgt gcacacctttt ccggctgtcc
6841 tacagtcctc aggactctac ttccttagca acgtggtgac cgtgccctcc agcagcttgg
6901 gcacccagac ctacatctgc aacgtgaatc acaagcccag caacaccaag gtggacaaga
6961 aagttgagcc caaatcttgt gacaaaactc acacatgccc accgtgccca gcacctgaac
7021 tcctgggggg accgtcagtc ttcctcttcc ccccaaaacc caaggacacc ctcatgatct
7081 cccggacccc tgaggtcaca tgcgtggtgg tggacgtgag ccacgaagac cctgaggtca
7141 agttcaactg gtacgtggac ggcgtggagg tgcataatgc caagacaaag ccgcgggagg
7201 agcagtacaa cagcacgtac cgtgtggtca gcgtcctcac cgtcctgcac caggactggc
7261 tgaatggcaa ggagtacaag tgcaaggtct ccaacaaagc cctcccagcc cccatcgaga
7321 aaaccatctc caaagccaaa gggcagcccc gagaaccaca ggtgtacacc ctgccccat
7381 cccgggatga gctgaccaag aaccaggtca gcctgacctg cctggtcaaa ggcttctatc
7441 ccagcgacat cgccgtggag tgggagagca atgggcagcc ggagaacaac tacaagacca
7501 cgcctcccgt gctggactcc gacggctcct tcttcctcta cagcaagctc accgtggaca
7561 agagcaggtg gcagcagggg aacgtcttct catgctccgt gatgcatgag gctctgcaca
7621 accactacac gcagaagagc ctctccctgt ctccgggtaa agcgccagag ccgaaaaagc
7681 ttcctatga gctgacacag ccaccctcgg tgtcagtgtc cccaggacaa acggccagga
7741 tcacctgctc tggagatgca ttgccagaaa aatatgttta ttggtaccag cagaagtcag
7801 gccaggcccc tgtggtggtc atctatgagg acagcaaacg accctcgggg atcgctgaga
7861 gattcctctgg ctccagctca gggacaatgg ccaccttgac tatcagtggg gcccaggtgg
7921 aagatgaagg tgactactac tgttactcaa ctgacagcag tggttatcat agggaggtgt
7981 tcagcggagg gaccaagctg accgtcctag gtcagcccaa ggctgccccc tcggtcactc
8041 tgttcccacc ctcctctgag gagcttcaag ccaacaaggc cacactggtg tgtctcataa
8101 gtgactccta cccgggagcc gtgacagtgg cctggaaggc agatagcagc cccgtcaagg
8161 cgggagtgga gaccaccaca ccctccaaac aaagcaacaa caagtacgcg gccagcagct
8221 acctgagcct gacgcttgag cagtggaagt cccacaaaag ctacagctgc caggtcacgc
8281 atgaagggag caccgtggag aagacagtgg cccctgcaga atgttcaccg cggagggagg

8341 gaagggccct ttttgaaggg ggaggaaact tcgcgccatg actcctctcg tgccccccgc
8401 acggaacact gatgtgcaga gggccctctg ccattgctgc ttcctctgcc cttcctcgtc
8461 actctgaatg tggcttcttt gctactgcca cagcaagaaa taaaatcttca acatctaaat
8521 gggttccctg agattttcca agagtcgtta agcacattcc ttccccagca cccctgctg
8581 caggccagtg ccaggcacca acttggctac tgctgcccat gagagaaatc cagttcaata
8641 ttttccaaag caaaatggat tacatatgcc ctagatcctg attaacaggt gtttgtatt
8701 atctgtgctt tcgcttcacc cacattatcc cattgcctcc cctcgagggg gggcccggta
8761 cccaattcgc cctatagtga gtcgtattac gcgcgctcac tggccgtcgt tttacaacgt
8821 cgtgactggg aaaaccctgg cgttacccaa cttaatcgcc ttgcagcaca tcccccttc
8881 gccagctggc gtaatagcga agaggcccgc accgatcgcc cttcccaaca gttgcgcagc
8941 ctgaatggcg aatggaaatt gtaagcgtta atatttgtt aaaattcgcg ttaaatttt
9001 gttaaatcag ctcattttt aaccaatagg ccgaaatcgg caaaatccct tataaatcaa
9061 aagaatagac cgagataggg ttgagtgttg ttccagtttg gaacaagagt ccactattaa
9121 agaacgtgga ctccaacgtc aaagggcgaa aaaccg-cta tcagggcgat ggcccactac
9181 tccgggatca tatgacaaga tgtgtatcca ccttaactta atgattttta ccaaaatcat
9241 taggggattc atcagtgctc agggtcaacg agaattaaca ttccgtcagg aaagcttatg
9301 atgatgatgt gcttaaaaac ttactcaatg gctggttatg catatcgcaa tacatgcgaa
9361 aaacctaaaa gagcttgccg ataaaaaagg ccaatttatt gctatttacc gcggcttttt
9421 attgagcttg aaagataaat aaaatagata ggttttattt gaagctaaat cttcttttatc
9481 gtaaaaaatg ccctcttggg ttatcaagag ggtcattata tttcgcggaa taacatcatt
9541 tggtgacgaa ataactaagc acttgtctcc tgtttactcc cctgagcttg agggttaac
9601 atgaaggtca tcgatagcag gataataata cagtaaaacg ctaaaccaat aatccaaatc
9661 cagccatccc aaattggtag tgaatgatta taaataacag caaacagtaa tgggccaata
9721 acaccggttg cattggtaag gctcaccaat aatcccctgta aagcaccttg ctgatgactc
9781 tttgtttgga tagacatcac tccctgtaat gcaggtaaag cgatcccacc accagcaat
9841 aaaattaaaa cagggaaaac taaccaacct tcagatataa acgctaaaaa ggcaaatgca
9901 ctactatctg caataaatcc gagcagtact gccgttttt cgcccattta gtggctattc
9961 ttcctgccac aaaggcttgg aatactgagt gtaaaagacc aagacccgta atgaaaagcc
10021 aaccatcatg ctattcatca tcacgatttc tgtaatagca ccacaccgtg ctggattggc
10081 tatcaatgcg ctgaaataat aatcaacaaa tggcatcgtt aaataagtga tgtataccga
10141 tcagctttg ttccctttag tgagggttaa ttgcgcgctt ggcgtaatca tggtcataagc
10201 tgtttcctgt gtgaaattgt tatccgctca caattccaca caacatacga gccggaagca
10261 taaagtgtaa agcctggggt gcctaatgag tgagctaact cacattaatt gcgttgcgct
10321 cactgcccgc tttccagtcg ggaaacctgt cgtgccagct gcattaatga atcggccaac
10381 gcgcggggag aggcggtttg cgtattgggc gctcttccgc ttcctcgctc actgactcgc
10441 tgcgctcggt cgttcggctg cggcgagcgg tatcagctca ctcaaaggcg gtaatacggt
10501 tatccacaga atcaggggat aacgcaggaa agaacatgtg agcaaaaggc cagcaaaagg
10561 ccaggaaccg taaaaaggcc gcgttgctgg cgtttttcca taggctccgc ccccctgacg
10621 agcatcacaa aatcgacgc tcaagtcaga ggtggcgaaa cccgacagga ctataaagat

```
10681 accaggcgtt tccccctgga agctccctcg tgcgctctcc tgttccgacc ctgccgctta
10741 ccggatacct gtccgccttt ctcccttcgg gaagcgtggc gctttctcat agctcacgct
10801 gtaggtatct cagttcggtg taggtcgttc gctccaagct gggctgtgtg cacgaacccc
10861 ccgttcagcc cgaccgctgc gccttatccg gtaactatcg tcttgagtcc aacccggtaa
10921 gacacgactt atcgccactg gcagcagcca ctggtaacag gattagcaga gcgaggtatg
10981 taggcggtgc tacagagttc ttgaagtggt ggcctaacta cggctacact agaaggacag
11041 tatttggtat ctgcgctctg ctgaagccag ttaccttcgg aaaaagagtt ggtagctctt
11101 gatccggcaa acaaaccacc gctggtagcg gtggtttttt tgtttgcaag cagcagatta
11161 cgcgcagaaa aaaggatct caagaagatc ctttgatctt ttctacgggg tctgacgctc
11221 agtggaacga aaactcacgt taagggattt tggtcatgag attatcaaaa aggatcttca
11281 cctagatcct ttaaattaa aaatgaagtt ttaaatcaat ctaaagtata tatgagtaaa
11341 cttggtctga cagttaccaa tgcttaatca gtgaggcacc tatctcagcg atctgtctat
11401 ttcgttcatc catagttgcc tgactccccg tcgtgtagat aactacgata cgggagggct
11461 taccatctgg ccccagtgct gcaatgatac cgcgagaccc acgctcaccg gctccagatt
11521 tatcagcaat aaaccagcca gccggaaggg ccgagcgcag aagtggtcct gcaactttat
11581 ccgcctccat ccagtctatt aattgttgcc gggaagctag agtaagtagt tcgccagtta
11641 atagtttgcg caacgttgtt gccattgcta caggcatcgt ggtgtcacgc tcgtcgtttg
11701 gtatggcttc attcagctcc ggttcccaac gatcaaggcg agttacatga tcccccatgt
11761 tgtgcaaaaa agcggttagc tccttcggtc ctccgatcgt tgtcagaagt aagttggccg
11821 cagtgttatc actcatggtt atggcagcac tgcataattc tcttactgtc atgccatccg
11881 taagatgctt ttctgtgact ggtgagtact caaccaagtc attctgagaa tagtgtatgc
11941 ggcgaccgag ttgctcttgc ccggcgtcaa tacgggataa tacgcgccca catagcagaa
12001 ctttaaaagt gctcatcatt ggaaaacgtt cttcggggcg aaaactctca aggatcttac
12061 cgctgttgag atccagttcg atgtaaccca ctcgtgcacc caactgatct tcagcatctt
12121 ttactttcac cagcgtttct gggtgagcaa aaacaggaag gcaaaatgcc gcaaaaaagg
12181 gaataagggc gacacggaaa tgttgaatac tcatactctt cctttttcaa tattattgaa
12241 gcatttatca gggttattgt ctcatgagcg gatacatatt tgaatgtatt tagaaaaata
12301 aacaaataggg ggttccgcgc acatttcccc gaaaagtgcc ac

SEQ ID NO:104 (pTnMOD (CMV-prepro-HCPro-CPA))
    1 ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga
   61 ccgctacact tgccagcgcc ctagcgcccg ctcctttcgc tttcttccct tcctttctcg
  121 ccacgttcgc cggcatcaga ttggctattg gccattgcat acgttgtatc catatcataa
  181 tatgtacatt tatattggct catgtccaac attaccgcca tgttgacatt gattattgac
  241 tagttattaa tagtaatcaa ttacgggggtc attagttcat agcccatata tggagttccg
  301 cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt
  361 gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca
  421 atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc
  481 aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta
```

```
 541 catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac
 601 catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg
 661 atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg
 721 ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt
 781 acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg
 841 ccatccacgc tgttttgacc tccatagaag acaccgggac cgatccagcc tccgcggccg
 901 ggaacggtgc attggaacgc ggattccccg tgccaagagt gacgtaagta ccgcctatag
 961 actctatagg cacacccctt tggctcttat gcatgctata ctgttttttg cttggggcct
1021 atacacccgc gcttccttat gctataggtg atggtatagc ttagcctata ggtgtgggtt
1081 attgaccatt atgaccact ccctattggg tgacgatact ttccattact aatccataac
1141 atggctcttt gccacaacta tctctattgg ctatatgcca atactctgtc cttcagagac
1201 tgacacggac tctgtatttt tacaggatgg ggtcccattt attatttaca aattcacata
1261 tacaacaacg ccgtcccccg tgcccgcagt ttttattaaa catagcgtgg gatctccacg
1321 cgaatcttcgg gtacgtgttc cggacatggg ctcttc*ccg gtagcggcgg agcttccaca
1381 tccgagccct ggtcccatgc ctccagcggc tcatgytcgc tcggcagctc cttgctccta
1441 acagtggagg ccagacttag gcacagcaca atgcccacca ccaccagtgt gccgcacaag
1501 gccgtggcgg tagggtatgt gtctgaaaat gagcgtggag attgggctcg cacggctgac
1561 gcagatggaa gacttaaggc agcggcagaa gaagatgcag gcagctgagt tgttgtattc
1621 tgataagagt cagaggtaac tcccgttgcg gtgctgttaa cggtggaggg cagtgtagtc
1681 tgagcagtac tcgttgctgc cgcgcgcgcc accagacata atagctgaca gactaacaga
1741 ctgttccttt ccatgggtct tttctgcagt caccgtcgga ccatgtgtga acttgatatt
1801 ttacatgatt ctctttacca atcctgcccc gaattacact taaaacgact caacagctta
1861 acgttggctt gccacgcatt acttgactgt aaaactctca ctctttaccga acttggccgt
1921 aacctgccaa ccaaagcgag aacaaaacat aacatcaaac gaatcgaccg attgttaggt
1981 aatcgtcacc tccacaaaga gcgactcgct gtataccgtt ggcatgctag ctttatctgt
2041 tcgggcaata cgatgcccat tgtacttgtt gactggtctg atattcgtga gcaaaaacga
2101 cttatggtat tgcgagcttc agtcgcacta cacggtcgtt ctgttactct ttatgagaaa
2161 gcgttcccgc tttcagagca atgttcaaag aaagctcatg accaatttct agccgacctt
2221 gcgagcattc taccgagtaa caccacaccg ctcattgtca gtgatgctgg ctttaaagtg
2281 ccatggtata aatcgcttga gaagctgggt tggtactggt taagtcgagt aagaggaaaa
2341 gtacaatatg cagacctagg agcggaaaac tggaaaccta tcagcaactt acatgatatg
2401 tcatctagtc actcaaagac tttaggctat aagaggctga ctaaaagcaa tccaatctca
2461 tgccaaattc tattgtataa atctcgctct aaaggccgaa aaaatcagcg ctcgacacgg
2521 actcattgtc accacccgtc acctaaaatc tactcagcgt cggcaaagga gccatgggtt
2581 ctagcaacta acttacctgt tgaaattcga acacccaaac aacttgttaa tatcttattcg
2641 aagcgaatgc agattgaaga aaccttccga gacttgaaaa gtcctgccta cggactaggc
2701 ctacgccata gccgaacgag cagctcagag cgttttgata tcatgctgct aatcgccctg
2761 atgcttcaac taacatgttg gcttgcgggc gttcatgctc agaaacaagg ttgggacaag
2821 cacttccagg ctaacacagt cagaaatcga aacgtactct caacagttcg cttaggcatg
```

```
2881 gaagttttgc ggcattctgg ctacacaata acaagggaag acttactcgt ggctgcaacc
2941 ctactagctc aaaatttatt cacacatggt tacgctttgg ggaaattatg ataatgatcc
3001 agatcacttc tggctaataa aagatcagag ctctagagat ctgtgtgttg gttttttgtg
3061 gatctgctgt gccttctagt tgccagccat ctgttgtttg ccctcccccc gtgccttcct
3121 tgaccctgga aggtgccact cccactgtcc tttcctaata aaatgaggaa attgcatcgc
3181 attgtctgag taggtgtcat tctattctgg ggggtggggt ggggcagcac agcaaggggg
3241 aggattggga agacaatagc aggcatgctg gggatgcggt gggctctatg ggtacctctc
3301 tctctctctc tctctctctc tctctctctc tctctcggta cctctctctc tctctctctc
3361 tctctctctc tctctctctc tcggtaccag gtgctgaaga attgaccggg tgaccaaagg
3421 tgccttttat catcacttta aaaataaaaa acaattactc agtgcctgtt ataagcagca
3481 attaattatg attgatgcct acatcacaac aaaaactgat ttaacaaatg gttggtctgc
3541 cttagaaagt atatttgaac attatcttga ttatattatt gataataata aaaaccttat
3601 ccctatccaa gaagtgatgc ctatcattgg ttggaatgaa cttgaaaaaa attagccttg
3661 aatacattac tggtaaggta aacgccattg tcagcaaatt gatccaagag aaccaactta
3721 aagctttcct gacggaatgt taattctcgt tgaccctgag cactgatgaa tcccctaatg
3781 atttggtaa aaatcattaa gttaaggtgg atacacatct tgtcatatga tcccggtaat
3841 gtgagttagc tcactcatta ggcacccaag gctttacact ttatgcttcc ggctcgtatg
3901 ttgtgtggaa ttgtgagcgg ataacaattt cacacaggaa acagctatga ccatgattac
3961 gccaagcgcg caattaaccc tcactaaagg gaacaaaagc tggagctcca ccgcggtggc
4021 ggccgctcta gaactagtgg atcccccggg ctgcaggaat tcgatatcaa gcttatcgat
4081 accgctgacc tcgagcatca gattggctat tggccattgc atacgttgta tccatatcat
4141 aatatgtaca tttatattgg ctcatgtcca acattaccgc catgttgaca ttgattattg
4201 actagttatt aatagtaatc aattacgggg tcattagttc atagcccata tatggagttc
4261 cgcgttacat aacttacggt aaatggcccg cctggctgac cgcccaacga cccccgccca
4321 ttgacgtcaa taatgacgta tgttcccata gtaacgccaa tagggacttt ccattgacgt
4381 caatgggtgg agtatttacg gtaaactgcc cacttggcag tacatcaagt gtatcatatg
4441 tcaagtacgc cccctattga cgtcaatgac ggtaaatggc ccgcctggca ttatgcccag
4501 tacatgacct tatgggactt tcctacttgg cagtacatct acgtattagt catcgctatt
4561 accatggtga tgcggttttg gcagtacatc aatgggcgtg gatagcggtt tgactcacgg
4621 ggatttccaa gtcttcaccc cattgacgtc aatgggagtt tgttttggca ccaaaatcaa
4681 cgggactttc caaaatgtcg taacaactcc gcccattga cgcaaatggg cggtaggcgt
4741 gtacggtggg aggtctatat aagcagagct cgtttagtga accgtcagat cgcctggaga
4801 cgccatccac gctgtttga cctccataga agacaccggg accgatccag cctccgcggc
4861 cgggaacggt gcattggaac gcggattccc cgtgccaaga gtgacgtaag taccgcctat
4921 agactctata ggcacacccc tttggctctt atgcatgcta tactgtttct ggcttggggc
4981 ctatacaccc ccgcttcctt atgctatagg tgatggtata gcttagccta taggtgtggg
5041 ttattgacca ttattgacca ctccctatt ggtgacgata ctttccatta ctaatccata
5101 acatggctct ttgccacaac tatctctatt ggctatatgc caatactctg tccttcagag
5161 actgacacgg actctgtatt tttacaggat ggggtcccat ttattattta caaattcaca
```

5221 tatacaacaa cgccgtcccc cgtgcccgca gttttatta aacatagcgt gggatctcca
5281 cgcgaatctc gggtacgtgt tccggacatg ggctcttctc cggtagcggc ggagcttcca
5341 catccgagcc ctggtcccat gcctccagcg gctcatggtc gctcggcagc tccttgctcc
5401 taacagtgga ggccagactt aggcacagca caatgcccac caccaccagt gtgccgcaca
5461 aggccgtggc ggtagggtat gtgtctgaaa atgagcgtgg agattgggct cgcacgggctg
5521 acgcagatgg aagacttaag gcagcggcag aagaagatgc aggcagctga gttgttgtat
5581 tctgataaga gtcagaggta actcccgttg cggtgctgtt aacggtggag ggcagtgtag
5641 tctgagcagt actcgttgct gccgcgcgcg ccaccagaca taatagctga cagactaaca
5701 gactgttcct ttccatgggt cttttctgca gtcaccgtcg gatcaatcat tcatctcgtg
5761 acttcttcgt gtgtggtgtt tacctatata tcttaaattta atatttcgtt tattaaaatt
5821 taatatattt cgacgatgaa tttctcaagg atatttttct tcgtgttcgc tttggttctg
5881 gctttgtcaa cagtttcggc tgcgccagag ccgaaaggta cccaggtgca gctgcaggag
5941 tcgggggagg gcttggtaaa gccgggggcg tcccttagag tctcctgtgc agcctctgga
6001 ttcactttca gaaacgcctg gatgagctgg gtccgcragg ctccagggaa ggggctggag
6061 tgggtcggcc gtattaaaag caaaattgat ggtgggacaa cagactatgc tgcacccgtg
6121 aaaggcagat tcaccatctc aagagatgat tcaaaaaaca cgttatatct gcaaatgaat
6181 agcctgaaag ccgaggacac agccgtatat tactgtacca cggggattat gataacattt
6241 ggggagtta tccctccccc gaattggggc cagggaaccc tggtcaccgt ctcctcagcc
6301 tccaccaagg gcccatcggt cttccccctg gcaccctcct ccaagagcac ctctgggggc
6361 acagcggccc tgggctgcct ggtcaaggac tacttccccg aaccggtgac ggtgtcgtgg
6421 aactcaggcg ccctgaccag cggcgtgcac acctttccgg ctgtcctaca gtccttcagga
6481 ctctacttcc ttagcaacgt ggtgaccgtg ccctccagca gcttgggcac ccagacctac
6541 atctgcaacg tgaatcacaa gcccagcaac accaaggtgg acaagaaagt tgagcccaaa
6601 tcttgtgaca aaactcacac atgcccaccg tgcccagcac ctgaactcct gggggaccg
6661 tcagtcttcc tcttcccccc aaaacccaag gacacctca tgatctcccg gacccctgag
6721 gtcacatgcg tggtggtgga cgtgagccac gaagaccctg aggtcaagtt caactggtac
6781 gtggacggcg tggaggtgca taatgccaag acaaagccgc gggaggagca gtacaacagc
6841 acgtaccgtg tggtcagcgt cctcaccgtc ctgcaccagg actggctgaa tggcaaggag
6901 tacaagtgca aggtctccaa caaagccctc ccagcccca tcgagaaaac catctccaaa
6961 gccaaagggc agccccgaga accacaggtg tacaccctgc ccccatcccg ggatgagctg
7021 accaagaacc aggtcagcct gacctgcctg gtcaaaggct tctatcccag cgacatcgcc
7081 gtggagtggg agagcaatgg gcagccggag aacaactaca agaccacgcc tccgtgctg
7141 gactccgacg gctccttctt cctctacagc aagctcaccg tggacaagag caggtggcag
7201 caggggaacg tcttctcatg ctccgtgatg catgaggctc tgcacaacca ctacacgcag
7261 aagagcctct ccctgtctcc gggtaaagcg ccagagccga agctttccta tgagctgaca
7321 cagccaccct cggtgtcagt gtcccaggaa caaacggcca ggatcacctg ctctggagat
7381 gcattgccag aaaaatatgt ttattggtac cagcagaagt caggccaggc ccctgtggtg
7441 gtcatctatg aggacagcaa acgaccctcc gggatccctg agagattctc tggctccagc
7501 tcagggacaa tggccacctt gactatcagt ggggcccagg tggaagatga aggtgactac

```
7561 tactgttact caactgacag cagtggttat cataggagg tgttcagcgg agggaccaag
7621 ctgaccgtcc taggtcagcc caaggctgcc ccctcggtca ctctgttccc accctcctct
7681 gaggagcttc aagccaacaa ggccacactg gtgtgtctca taagtgactc ctacccggga
7741 gccgtgacag tggcctggaa ggcagatagc agccccgtca aggcgggagt ggagaccacc
7801 acaccctcca aacaaagcaa caacaagtac gcggccagca gctacctgag cctgacgctt
7861 gagcagtgga agtcccacaa aagctacagc tgccaggtca cgcatgaagg gagcaccgtg
7921 gagaagacag tggcccctgc agaatgttca ccgcggaggg agggaagggc ccttttgaa
7981 ggggaggaa acttcgcgcc atgactcctc tcgtgccccc cgcacggaac actgatgtgc
8041 agagggccct ctgccattgc tgcttcctct gccctcctga gtcactctga atgtggcttc
8101 tttgctactg ccacagcaag aaataaaatc tcaacatcta aatgggtttc ctgagatttt
8161 tcaagagtcg ttaagcacat tccttcccca gcacccttg ctgcaggcca gtgccaggca
8221 ccaacttggc tactgctgcc catgagaga atccagttca atattttcca aagcaaaatg
8281 gattacatat gccctagatc ctgattaaca ggtgttttgt attatctgtg ctttcgcttc
8341 acccacatta tcccattgcc tcccctcgac tcgaggggg gcccggtacc caattcgccc
8401 tatagtgagt cgtattacgc gcgctcactg gccgtcgttt tacaacgtcg tgactgggaa
8461 aaccctggcg ttacccaact taatcgcctt gcagcacatc ccccttcgc cagctggcgt
8521 aatagcgaag aggcccgcac cgatcgccct tcccaacagt tgcgcagcct gaatggcgaa
8581 tggaaattgt aagcgttaat atttgttaa aattcgcgtt aaatttttgt taaatcagct
8641 catttttaa ccaataggcc gaaatcggca aaatcccttа taaatcaaaa gaatagaccg
8701 agatagggtt gagtgttgtt ccagtttgga acaagagtcc actattaaag aacgtggact
8761 ccaacgtcaa agggcgaaaa accgtcttatc agggcgatgg cccactactc cgggatcata
8821 tgacaagatg tgtatccacc ttaacttaat gatttttacc aaaatcatta ggggattcat
8881 cagtgctcag ggtcaacgag aattaacatt ccgtcaggaa agcttatgat gatgatgtgc
8941 ttaaaaactt actcaatggc tggttatgca tatcgcaata catgcgaaaa acctaaaaga
9001 gcttgccgat aaaaaggcc aatttattgc tatttaccgc ggcttttat tgagcttgaa
9061 agataaataa aatagatagg ttttatttga agctaaatct tctttatcgt aaaaaatgcc
9121 ctcttgggtt atcaagaggg tcattatatt tcgcggaata acatcatttg gtgacgaaat
9181 aactaagcac ttgtctcctg tttactcccc tgagcttgag gggttaacat gaaggtcatc
9241 gatagcagga taataataca gtaaaacgct aaaccaataa tccaaaatcca gccatcccaa
9301 attggtagtg aatgattata aataacagca aacagtaatg ggccaataac accggttgca
9361 ttggtaaggc tcaccaataa tccctgtaaa gcaccttgct gatgactctt tgtttggata
9421 gacatcactc cctgtaatgc aggtaaagcg atcccaccac cagccaataa aattaaaaca
9481 gggaaaacta accaaccttc agatataaac gctaaaaagg caaatgcact actatctgca
9541 ataaatccga gcagtactgc cgttttttcg cccatttagt ggctattctt cctgccacaa
9601 aggcttggaa tactgagtgt aaaagaccaa gacccgtaat gaaaagccaa ccatcatgct
9661 attcatcatc acgattcctg taatagcacc acaccgtgct ggattggcta tcaatgcgct
9721 gaaataataa tcaacaaatg gcatcgttaa ataagtgatg tataccgatc agctttgttt
9781 ccctttagtg agggttaatt gcgcgcttgg cgtaatcatg gtcatagctg tttcctgtgt
9841 gaaattgtta tccgctcaca attccacaca acatacgagc cggaagcata aagtgtaaag
```

```
 9901 cctggggtgc ctaatgagtg agctaactca cattaattgc gttgcgctca ctgccgcttt
 9961 tccagtcggg aaacctgtcg tgccagctgc attaatgaat cggccaacgc gcggggagag
10021 gcggtttgcg tattgggcgc tcttccgctt cctcgctcac tgactcgctg cgctcggtcg
10081 ttcggctgcg gcgagcggta tcagctcact caaaggcggt aatacggttа tccacagaat
10141 caggggataa cgcaggaaag aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta
10201 aaaaggccgc gttgctggcg tttttccata ggctccgccc ccctgacgag catcacaaaa
10261 atcgacgctc aagtcagagg tggcgaaacc cgacaggact ataaagatac caggcgtttc
10321 ccccctggaag ctccctcgtg cgctctcctg ttccgaccct gccgcttacc ggatacctgt
10381 ccgcctttct cccttcgggа agcgtggcgc tttctcatag ctcacgctgt aggtatctca
10441 gttcggtgta ggtcgttcgc tccaagctgg gctgtgtgca cgaacccccc gttcagcccg
10501 accgctgcgc cttatccggt aactatcgtc ttgagtccaa cccggtaaga cacgacttat
10561 cgccactggc agcagccact ggtaacagga ttagcagagc gaggtatgta ggcggtgcta
10621 cagagttctt gaagtggtgg cctaactacg gctcactag aaggacagta tttggtatct
10681 gcgctctgct gaagccagtt accttcggaa aaagagrtgg tagctcttga tccggcaaac
10741 aaaccaccgc tggtagcggt ggtttttttg ttgcaagca gcagattacg cgcagaaaaa
10801 aaggatctca agaagatcct ttgatctttt ctacgggggtc tgacgctcag tggaacgaaa
10861 actcacgtta agggattttg gtcatgagat tatcaaaaag gatcttcacc tagatccttt
10921 taaattaaaa atgaagtttt aaatcaatct aaagtatata tgagtaaact tggtctgaca
10981 gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctattt cgttcatcca
11041 tagttgcctg actccccgtc gtgtagataa ctacgatacg ggagggctta ccatctggcc
11101 ccagtgctgc aatgataccg cgagacccac gctcaccggc tccagattta tcagcaataa
11161 accagccagc cggaagggcc gagcgcagaa gtggtcctgc aactttatcc gcctccatcc
11221 agtctattaa ttgttgccgg gaagctagag taagtagttc gccagttaat agtttgcgca
11281 acgttgttgc cattgctaca ggcatcgtgg tgtcacgctc gtcgtttggt atggcttcat
11341 tcagctccgg ttcccaacga tcaaggcgag ttacatgatc ccccatgttg tgcaaaaaag
11401 cggttagctc cttcggtcct ccgatcgttg tcagaagtaa gttggccgca gtgttatcac
11461 tcatggttat ggcagcactg cataattctc ttactgtcat gccatccgta agatgctttt
11521 ctgtgactgg tgagtactca accaagtcat tctgagaata gtgtatgcgg cgaccgagtt
11581 gctcttgccc ggcgtcaata cgggataata ccgcgccaca tagcagaact ttaaaagtgc
11641 tcatcattgg aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat
11701 ccagttcgat gtaacccact cgtgcaccca actgatcttc agcatctttt actttcacca
11761 gcgtttctgg gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga ataagggcga
11821 cacggaaatg ttgaatactc atactcttcc tttttcaata ttattgaagc atttatcagg
11881 gttattgtct catgagcgga tacatatttg aatgtattta gaaaaataaa caaatagggg
11941 ttccgcgcac atttccccga aaagtgccac
```

CLAIMS

We Claim:

1. An isolated polynucleotide comprising at least one gene of interest and one or more pro nucleotide sequences, wherein the at least one gene of interest is operably-linked to a pro nucleotide sequence and each of the at least one gene of interest may be the same or different when more than one one gene of interest is present, wherein the at least one gene of interest encodes a protein useful in a vaccine.

2. The polynucleotide of claim 1, wherein a most 5' pro nucleotide sequence of the one or more pro nucleotide sequences is a part of a prepro nucleotide sequence.

3. The polynucleotide of claim 1, wherein two genes of interest and two pro nucleotide sequences are arranged in the following order: a prepro nucleotide sequence, a first gene of interest, a pro nucleotide sequence, and a second gene of interest.

4. The polynucleotide of claim 3, wherein the prepro nucleotide sequence is a cecropin prepro nucleotide sequence and the pro nucleotide sequence is a cecropin pro sequence.

5. The polynucleotide of claim 3, wherein the prepro nucleotide sequence comprises a sequence shown in SEQ ID NO:3 or SEQ ID NO:4 and the pro nucleotide sequence comprises a sequence shown in SEQ ID NO:1 or SEQ ID NO:2.

6. The polynucleotide of claim 1, wherein a first gene of interest encodes for an antibody heavy chain and a second gene of interest encodes for an antibody light chain.

7. A method of producing a multimeric protein in an individual comprising administering to the individual a polynucleotide comprising two or

more genes of interest, wherein each gene of interest encodes a part of the multimeric protein, each gene of interest is operably-linked to a pro nucleotide sequence, and each of the two or more genes of interest may be the same or different, wherein each of the two or more genes of interest encodes a protein useful in a vaccine.

8. The method of claim 7, wherein a most 5' pro nucleotide sequence of the two or more pro sequences is a part of a prepro nucleotide sequence.

9. The method of claim 8, wherein the polynucleotide comprises two genes of interest and two pro nucleotide sequences arranged in the following order: a prepro nucleotide sequence, a first gene of interest, a pro nucleotide sequence, and a second gene of interest.

10. The method of claim 7, wherein a first gene of interest encodes for an antibody heavy chain and a second gene of interest encodes for an antibody light chain.

11. A method of producing a multimeric protein in an individual comprising administering to the individual a polynucleotide comprising a cecropin prepro nucleotide sequence operably-linked to two or more genes of interest, each gene of interest encoding a part of the multimeric protein, wherein the multimeric protein is useful in a vaccine.

12. The method of claim 11, wherein a first gene of interest is an antibody heavy chain and a second gene of interest is an antibody light chain.

13. A method of producing a multimeric protein in an individual comprising administering to the individual a polynucleotide comprising two or more genes of interest, wherein each gene of interest encodes a part of the multimeric protein and wherein each gene of interest is operably linked to a gene encoding for a cleavage site, wherein the multimeric protein is useful in a vaccine.

14. The method of claim 13, wherein a transposon-based vector comprises the polynucleotide and further comprises a transposase gene operably linked to a first promoter and wherein;

a) the first promoter comprises a modified Kozak sequence comprising ACCATG;

b) the two or more genes of interest are each operably-linked to one or more additional promoters; and,

c) the two or more genes of interest and their operably-linked promoters are flanked by transposase insertion sequences recognized by a transposase encoded by the transposase gene.

15. The method of claim 13, wherein a transposon-based vector comprises the polynucleotide and further comprises a transposase gene operably linked to a first promoter and an avian optimized polyA sequence, and wherein;

a) the two or more genes of interest are each operably-linked to one or more additional promoters; and,

b) the two or more genes of interest and their operably-linked promoters are flanked by transposase insertion sequences recognized by a transposase encoded by the transposase gene.

16. An animal comprising the isolated polynucleotide of Claim 1.

17. The animal of Claim 16, wherein the animal is a bird.

18. An egg produced by the animal of Claim 16.

19. The egg of Claim 18, wherein the egg comprises a multimeric protein encoded by the isolated polynucleotide.

20. The animal of Claim 16, wherein the animal is a mammal.

21. Milk produced by the mammal of Claim 20.

22. The milk of Claim 21, wherein the milk comprises a multimeric protein encoded by the isolated polynucleotide.

23. A method of producing a multimeric protein comprising:

a) administering to an egg-laying animal a composition comprising the polynucleotide of Claim 1; and,

b) permitting the one or more genes of interest to be expressed into the multimeric protein, wherein the multimeric protein is useful in a vaccine.

24. The method of Claim 23, further comprising

a) collecting an egg from the egg-laying animal;

b) harvesting egg white comprising the multimeric protein; and,

c) purifying the multimeric protein.

25. The method of Claim 23, wherein the egg-laying animal is a bird.

26. A method of producing a multimeric protein comprising:

a) administering to an intramammary duct system of a mammal a composition comprising the polynucleotide of Claim 1, and,

b) permitting the one or more genes of interest to be expressed into the multimeric protein, wherein the multimeric protein is useful in a vaccine.

27. The method of Claim 26, further comprising

a) collecting milk from the mammal, wherein the milk comprises the multimeric protein; and

b) purifying the multimeric protein.

28. A method of vaccinating an individual against a disease comprising administering to the individual a composition comprising a protein, a multimeric protein or an antibody or a combination thereof, wherein the

composition is produced in the individual or in another individual made transgenic using transposon-based vectors.

29. Use of a composition comprising a protein, a multimeric protein or an antibody or a combination thereof, wherein the composition is produced in an individual made transgenic using transposon-based vectors, in the preparation of a medicament for use in a vaccine.

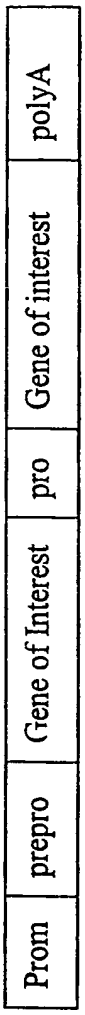
30. The polynucleotide of any of the preceding claims, wherein the prepro nucleotide sequence comprises a sequence shown in SEQ ID NO:3 or SEQ ID NO:4.

31. The polynucleotide of any of the preceding claims, wherein the two or more pro nucleotide sequences each comprise a sequence shown in SEQ ID NO:1 or SEQ ID NO:2.

32. The polynucleotide of any of the preceding claims, wherein the prepro nucleotide sequence is a cecropin prepro nucleotide sequence.

33. The polynucleotide of any of the preceding claims, wherein the pro nucleotide sequence is a cecropin pro nucleotide sequence.

FIGURE 1



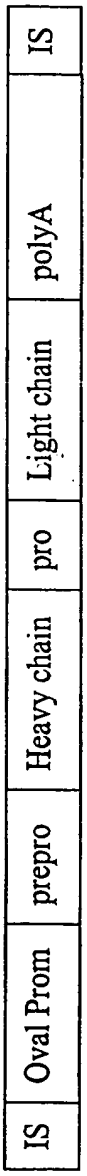


FIGURE 2

FIGURE 3


Prom
Cecropin prepro
Gene of Interest
CS
Gene of Interest
polyA

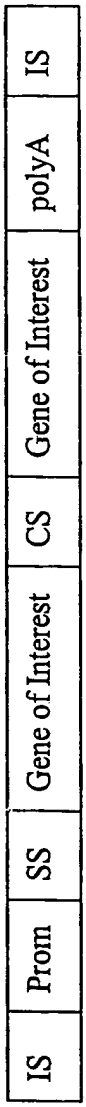


FIGURE 4

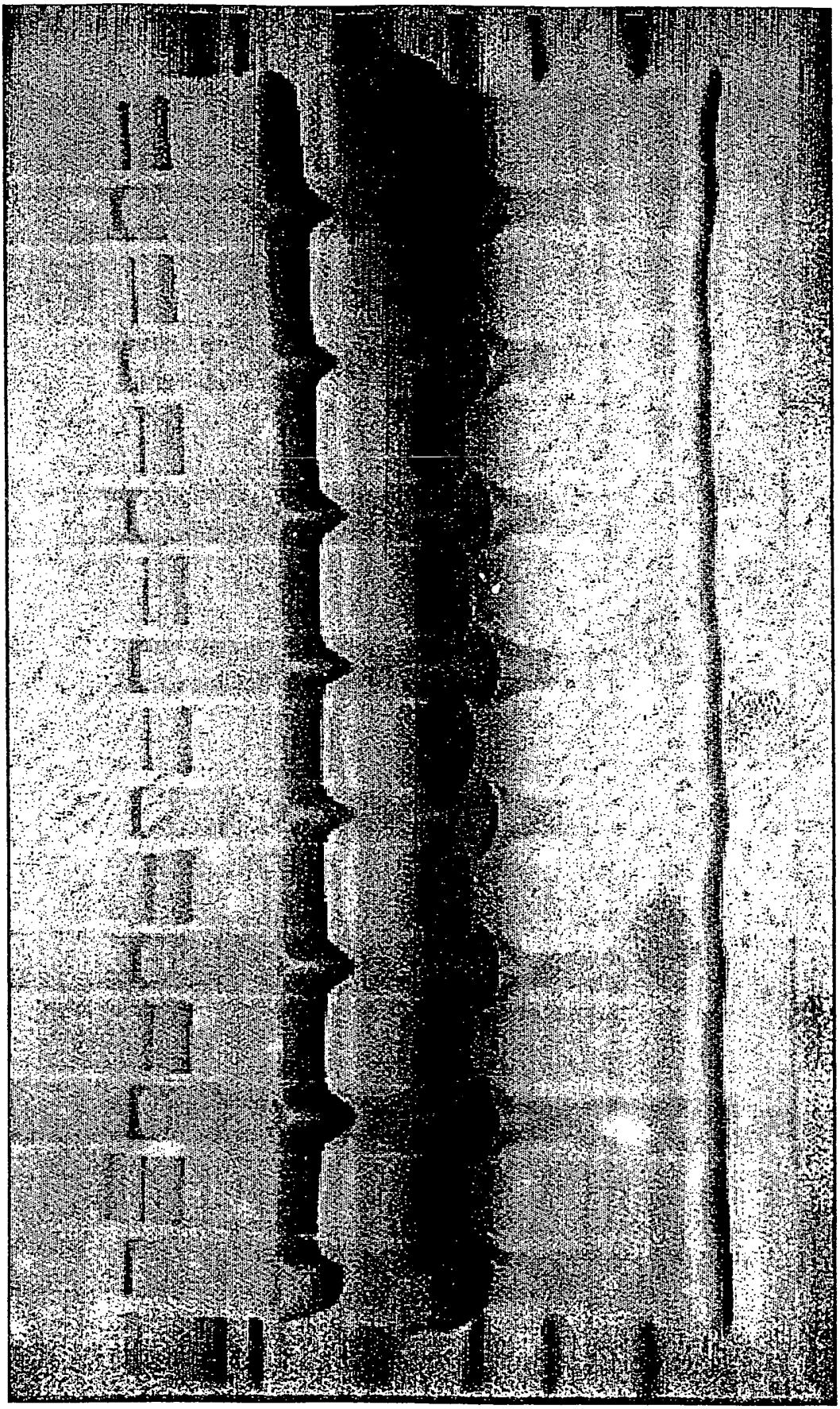


FIGURE 5

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.